

Janine Kaden

Energieprofilbasierende Analyse der Pfam Familie 2-
Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline
Decarboxylase

BACHELORARBEIT

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Mathematik, Naturwissenschaften, Informatik

Mittweida, 2011

Janine Kaden

Energieprofilbasierende Analyse der Pfam Familie 2-
Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline
Decarboxylase

eingereicht als

BACHELORARBEIT

an der

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Mathematik, Naturwissenschaften, Informatik

Mittweida, 2011

Erstprüfer: Prof. Dr. rer. nat. Dirk Labudde

Zweitprüfer: Dipl. – Inf. (FH) Daniel Stockmann

Bibliografische Beschreibung

Kaden, Janine

Energieprofilbasierende Analyse der Pfam Familie 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase. – 2011. – 90 S.

Mittweida, Hochschule Mittweida

Fakultät Mathematik, Naturwissenschaften, Informatik

Bachelorarbeit, 2011

Vorwort

Diese Arbeit ist durch eine gute Zusammenarbeit mit meinem Betreuer Herrn Prof. Dr. Dirk Labudde möglich geworden. Ich danke ihm für die beigesteuerten Ideen. Einen großen Dank gilt auch meinem Kommilitone Florian Heinke, welcher mir sehr bei der Ausführung der Energieprofilerstellung behilflich war.

Weiterhin danke ich meiner Familie, besonders meiner Mutter Petra und meinen Freund Rocco, für die Motivation und tatkräftige Unterstützung.

Kurzfassung

Im Folgenden wird die Enzymfamilie 2-oxo-4-Hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase beschrieben. Diese Enzymfamilie regelt die Umwandlung des Purinderivats 2-oxo-4-Hydroxy-4-carboxy-5-ureidoimidazoline zu dem sehr gut wasserlöslichen (S)-Allantoin im Purinkatabolismus. In der nachfolgenden Arbeit wird die Familie basierend auf Sequenz- und Energieprofilebene untersucht und beschrieben. Dabei wird das Ziel verfolgt herauszufinden, wie gut die Aussagen von Energieprofilen mit den Ergebnissen von Sequenzbasierenden Methoden korrelieren bzw. ob Energieprofilbasierende Analysen sogar genauere Resultate liefern können. Hierbei beispielsweise in Bezug auf strukturelle oder phylogenetische Untersuchungen. Zunächst werden im Punkt 1 theoretische Grundlagen für ein besseres Verständnis im Bezug auf Proteine bzw. Enzyme und auch für bioinformatische Tools und Werkzeuge vermittelt. Im Punkt 2 wird ein Einblick auf Wechselwirkungen und Energien von Aminosäuren in und zwischen Proteinen gegeben. Dabei wird die Berechnung der Energiefunktion von Aminosäuren, auf welcher aufbauend Energieprofile berechnet werden können, erläutert. Anschließend wird die 2-oxo-4-Hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase im Punkt 3 genauer vorgestellt. Hier steht die Strukturaufklärung des Enzyms und dessen Funktion im Purinmetabolismus im Vordergrund. Dem folgend finden im Punkt 4 Sequenz basierende Analysen der Familien statt. Neben der Erstellung eines multiplen Sequenz Alignments wird ein auf diesem Alignment rekonstruierter phylogenetischer Baum berechnet um erste phylogenetische Aussagen bzw. Eindrücke zu erzielen. Für Energieprofil basierende Untersuchungen bezüglich der Korrelation zwischen Energie, Konservierung und Sekundärstrukturelementen werden im Punkt 5 Energieprofile zweier bekannter Strukturen erzeugt. Des Weiteren wird ein multiples Energieprofil Alignment basierend auf Sequenzen ausgewählter Organismengruppen vorgestellt. Anhand der Energieprofile von Strukturen und Sequenzen wird anschließend ein phylogenetischer Baum rekonstruiert. Welcher im Punkt 6 mit dem auf Sequenzebene basierenden phylogenetischen Baum verglichen wird. Dabei gilt es zu untersuchen, ob sich Unterschiede herauskristallisieren und ob sich Gemeinsamkeiten feststellen lassen. Im abschließenden Punkt 7 wird ein Ausblick auf zukünftige Tätigkeiten bzw. Analysen der Familie gegeben.

Inhaltsverzeichnis

Abkürzungsverzeichnis.....	viii
Abbildungsverzeichnis.....	ix
Formelverzeichnis	xi
Tabellenverzeichnis.....	xii
 1. Theoretische Grundlagen	 1
1.1 Biologische Grundlagen	1
1.1.1 Proteine und Enzyme	1
1.1.2 Aufbau und Struktur der Proteine	5
1.1.3 Proteinfaltung	9
1.2 Bioinformatische Grundlagen und Werkzeuge.....	11
1.2.1 Protein Familien Datenbank	11
1.2.2 Multiples Sequenz Alignment	14
1.2.3 Phylogenetische Analyse	19
 2. Wechselwirkungen und Energiefunktion.....	 24
2.1 Intermolekulare Wechselwirkungen	24
2.2 Berechnung von Energieprofilen globulärer Proteine.....	28
 3. Analyse von 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase.....	 32
3.1 Strukturaufklärung und Modellierung des aktiven Zentrums	34
3.2 Funktion und Einordnung in den Purinkatabolismus	40
 4. Betrachtung der Familie OHCU Decarboxylase auf der Ebene der Sequenzen...	45
4.1 Multiple Alignments basierend auf Sequenzen	46
4.2 Erstellung eines phylogenetischen Baumes auf der Ebene der Sequenzen...	55
 5. Betrachtung der Familie OHCU Decarboxylase auf Ebene der Energieprofile	 58
5.1 Erstellung der Energieprofile basierend auf den Strukturen 2o70 und 2q37 ...	58
5.2 Erstellung eines Multiplen Energieprofil Alignments basierend auf den Energieprofilen von Sequenzen	63

5.3 Erstellung eines phylogenetischen Baumes auf Ebene der Energieprofile von Sequenzen.....	70
6 Vergleich der phylogenetischen Beobachtungen basierend auf Sequenz – und Energieprofilebene	72
7 Ausblick	74
Anhang	75
Literaturverzeichnis	77

Abkürzungsverzeichnis

AS	Aminosäure
ATP	Adenosintriphosphat
BLOSUM	Blocks Substitution Matrix
BSB	Biochemischer Sauerstoffbedarf
COOH	Carboxylgruppe
CSB	Chemischer Sauerstoffbedarf
DNS	Desoxyribonukleinsäure (eng. DNA)
EC ₀	(engl.) Effect Concentration
EC-Nr.	Enzyme Commission Nummer
EP	Energieprofil
H ₂ O ₂	Wasserstoffperoxid
HIU	5-Hydroxyisourate
HMM	Hidden Markov Model
LD	Letale Dosis
MEPAL	Multiples Energieprofil Alignment
MSA	Multiple Sequence Alignment
NH ₂	Aminogruppe
NJ	Neighbor Joining
OHCU	2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline
OHCU-Dec	2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase
PAM	Percent Accepted Mutation
PDB	Protein Data Bank
PEPAL	Paarweises Energieprofil Alignment
Pfam	Protein Families
PRODOM	Database of Protein Domain Families
PROSITE	Database of Protein Domains, Families and Functional Sites
PSSM	Position specific Scoring Matrix
SCOP	Structure Classification of Proteins
UPGMA	Unweight Pair-Group Method with Arithmetic Mean

Abbildungsverzeichnis

Abbildung 1 Reaktionsverlauf einer enzymatisch katalysierten- und einer nicht katalysierten Reaktion	1
Abbildung 2 1) Schloss-Schlüssel-Prinzip; 2) Induced-Fit-Modell.....	4
Abbildung 3 Venn Diagramm der Aminosäure Eigenschaften.....	5
Abbildung 4 Bildung einer Peptidbindung zwischen zwei Aminosäuren.....	6
Abbildung 5 Schema der dreidimensionalen Struktur der Peptidbindung.....	7
Abbildung 6 Dreidimensionale Darstellung des im Speichel vorkommenden Enzyms Amylase.....	8
Abbildung 7 Faltungsweg eines Polypeptides.	10
Abbildung 8 Die Pfam Familien Seite der Webseite.	11
Abbildung 9 Beispiel eines Hidden Markov Modells	12
Abbildung 10 ClustalW. Distanzmatrix von 7 Sequenzen.....	15
Abbildung 11 ClustalW: Ungewurzelter NJ-Baum.	16
Abbildung 12 ClustalW: Gewurzelter NJ-Baum.	16
Abbildung 13 Von Clustal W berechnetes MSA, Darstellung in JalView	17
Abbildung 14 Phylogenetischer Baum mit sechs OTUs und vier HTUs.....	20
Abbildung 15 Stammbäume. Rechts: Kladogramm; Links: Phylogramm.....	20
Abbildung 16 Dendrogramm nach UPGMA.	22
Abbildung 17 Möglichkeiten der Ausbildung von Haupt- und Nebervalenzbindungen bei Proteinmolekülen	24
Abbildung 18 Diagramm über die Verteilung bzw. Tendenz der Aminosäuren bestimmte Sekundärstrukturen zu bilden	27
Abbildung 19 Links: Ausschnitt des Energieprofil der Struktur 2Q37 Rechts: 3D-Struktur von 2Q37	31
Abbildung 20 Aminosäuresequenz der Zebrafisch OHCU Decarboxylase.	34
Abbildung 21 1) 3D-Darstellung des Dimers. 2) 3D-Darstellung von C α Kette des Dimers mit gebundenem Allantoin.....	37
Abbildung 22 Bildung der 2-Schritt-Achse	37
Abbildung 23 1) Darstellung der N-terminale Domäne eines Monomers. 2) Interaktionen 3) Wechselwirkungen zwischen gebundenem (S)-Allantoin und Resten des aktiven Zentrums.	39
Abbildung 24 Ausschnitt des Purinkatabolismus.	40

Abbildung 25 Vermutlicher Mechanismus der OHCU Decarboxylase	41
Abbildung 26 Blick auf den möglichen OHCU-OHCU Decarboxylase Komplex	42
Abbildung 27 Harnsäure Metabolismus.....	43
Abbildung 28 ClustalW Farbschema in JalView	46
Abbildung 29 JalView Ansicht des <i>Metazoa</i> -OHCU-Dec-MSA.....	46
Abbildung 30 JalView Ansicht des <i>Fungi</i> -OHCU-Dec-MSA.....	48
Abbildung 31 JalView Ansicht des <i>Actinobacteria</i> -OHCU-Dec-MSA.....	49
Abbildung 32 JalView Ansicht des <i>Proteobacteria</i> MSA.....	50
Abbildung 33 WebLogo (zwischen Position 68 und 218).....	53
Abbildung 34 Stark verkleinerter Phylogenetischer NJ Baum	55
Abbildung 35 EP der Struktur 2o70 (PDB-ID).....	59
Abbildung 36 EP der Struktur 2q37 (PDB-ID).....	61
Abbildung 37 Ausschnitt des PEPAL der zwei Strukturen 2o70 und 2o74.	63
Abbildung 38 EP-Alignment der zwei Strukturen 2o70 und 2q37	64
Abbildung 39 Paarweises Alignment von 2o70 und 2q37.....	65
Abbildung 40 ePlot von 2o70 und 2q37.....	66
Abbildung 41 3D-Darstellung der Strukturen 2o70 und 2q37	66
Abbildung 42 Multiples Energieprofil Alignment (MEPAL).	67
Abbildung 43 2o70 AS Sequenz mit SSE.....	68
Abbildung 44 MEPAL aus Struktur-Energieprofilen und Sequenz-Energieprofilen..	69
Abbildung 45 Phylogenetischer Baum.....	70
Abbildung 46 Vergleich der phylogenetischen Bäume der OHCU Decarboxylase basierend auf MSA und MEPAL.....	72
Abbildung 47 Phylogenetischer NJ Baum	76

Formelverzeichnis

Formel 1 Levinthal-Formel.....	9
Formel 2 Innen/Außen-Kriterium einer Aminosäure(i).	28
Formel 3 Boltzmannverteilung zur Berechnung der Lösungsenergie einer AS.....	29
Formel 4 Berechnung der Aminosäure-Aminosäure-Wechselwirkungsenergien.....	29
Formel 5 Berechnung der Energiedifferenz.	30
Formel 6 Berechnung des vereinfachten Paarpotentials von AS(i) und AS(j).....	30
Formel 7 Berechnung der Gesamtenergie eines globulären Proteins.	31

Tabellenverzeichnis

Tabelle 1	Tabellarische Übersicht des Dreibuchstabencodes und des Einbuchstabencodes der 20 Aminosäuren.....	7
Tabelle 2	Tabellarische Verteilung der Aminosäuren ein bestimmtes Sekundärstrukturelement zu bilden	26
Tabelle 3	Die Innen/Außen-Verteilung von Aminosäuren.....	28
Tabelle 4	Tabellarische Zusammenfassung der Strukturelemente.....	36
Tabelle 5	Tabellarische Darstellung der 100%-ig konservierten AS	51
Tabelle 6	Verteilung zwischen Energiequantile und Konservierungsgrad	60
Tabelle 7	Tabellarische Gegenüberstellung der Strukturen 2o70 und 2q37	61
Tabelle 8	Einordnung aller AS der Strukturen 2o70 und 2q37 in die entsprechenden Sekundärstrukturelemente (SSE) und Energiequantile zur Erfassung der Energienverteilung in den SSE.....	62
Tabelle 9	Überblick über Chemisch-Physikalische-, toxikologische- und ökotoxikologische Eigenschaften von Allantoin	75

1. Theoretische Grundlagen

1.1 Biologische Grundlagen

1.1.1 Proteine und Enzyme

Proteine sind dreidimensionale dynamische Makromoleküle und bestehen aus 100 bis 10000 Atomen. Meist weisen sie eine kompakte Struktur auf. Kommen Proteine im Cytoplasma bzw. in entsprechenden Zellorganellen vor so handelt es sich um globuläre Proteine. Proteine, welche hingegen in der Zellhülle sitzen, werden Membranproteine genannt. Die Zusammensetzung und die Konzentration der Proteine variiert je nach Lebensbedingungen der Zelle. Ein komplexes Netzwerk sorgt für die Regulation der Proteinsynthese, sodass nicht alle Proteine gleichzeitig auf einem hohen Niveau sondern bedarfsorientiert produziert werden. Somit geht auch keine wertvolle Energie verloren [1][2].

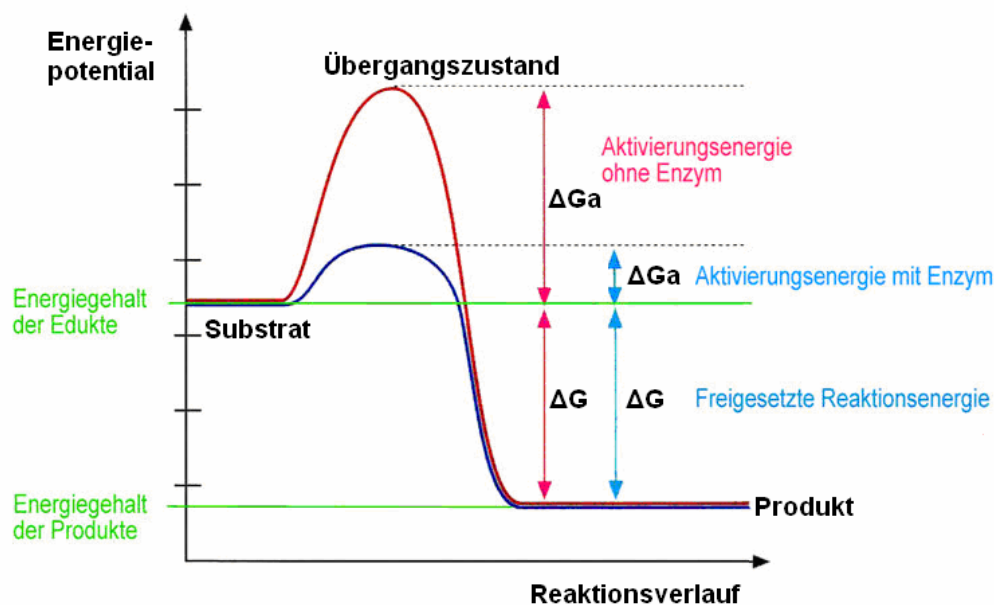


Abbildung 1 Reaktionsverlauf einer enzymatisch katalysierten- (blaue Kurve) und einer nicht katalysierten Reaktion (rote Kurve). Die grünen Linien entsprechen dem Energiegehalt der Edukte bzw. Produkte. Der obere rosa bzw. obere blaue Pfeil ΔG_a stellt die Aktivierungsenergie der Reaktion ohne bzw. mit Enzym dar, der untere rosa (ohne Enzym) bzw. blaue (mit Enzym) Pfeil ΔG stellt die freigesetzte Reaktionsenergie dar. [Quelle: <http://www.edyounet.de>]

Proteine können neben vielen Funktionen, beispielsweise Transportfunktionen in den Membranen, auch als Enzym, d.h. als Biokatalysator wirken. Enzyme sind

Biokatalysatoren, da sie eine biologische Reaktion katalysieren, bzw. überhaupt zum laufen bringen und somit auch beschleunigen (siehe Abbildung 1). Nach der Reaktion liegen sie in ihrer ursprünglichen Form vor.

Wie in Abbildung 1 dargestellt, wird eine Triebkraft benötigt um eine Reaktion zu starten. Diese Triebkraft ist die Änderung der freien Enthalpie ΔG , welche sich aus der Enthalpieänderung ΔH (Wärme, die bei der Reaktion umgesetzt wird) und der Änderung der Entropie ΔS (Ordnungszustand) zusammensetzt. Daraus lässt sich folgende Formel ableiten: $\Delta G = \Delta H - T\Delta S$, wobei T die Temperatur angibt. ΔG gibt die Energie an, die bei der Umsetzung freigesetzt wird oder aufgewendet werden muss und hängt dabei von der molaren Konzentration des Substrates und des Produktes ab. Zusätzliches hängt ΔG von einer für jede Reaktion typischen Konstante ΔG° , welche die Änderungen der Standardenthalpie bei pH 7,0, ab. Wenn ΔG einen negativen Wert erreicht, so läuft die Reaktion freiwillig, meist aber sehr langsam an. Die Reaktionsgeschwindigkeit hängt aber nicht von ΔG sondern von der Aktivierungsenergie ΔG_a ab. Diese muss das Substrat aufnehmen, um den Scheitelpunkt der Energiepotential-Reaktionsverlauf-Kurve (Abbildung 1) zu erreichen. An diesem Scheitelpunkt befinden sich die Reaktanten im Überangebot. Das Enzym lässt die Reaktion über andere Wege und Übergangszustände mit niedrigerer Aktivierungsenergie verlaufen und beschleunigt so die Reaktion. Somit ändert das Enzym nicht die freie Enthalpie ΔG sondern die Reaktionsgeschwindigkeit [3]. Alles in allem ist die Aktivierungsenergie ΔG_a (Abbildung 1, oberer blauer Pfeil) einer enzymatisch katalysierten Reaktion (Abbildung 1, blaue Kurve) niedriger als die Aktivierungsenergie ΔG_a (Abbildung 1, oberer rosa Pfeil) einer nicht katalysierten Reaktion (Abbildung 1, rote Kurve). Doch die Menge der freigesetzten Energie ΔG , welche während der katalysierten (Abbildung 1, unterer blauer Pfeil) und der nicht katalysierten (Abbildung 1, unterer rosa Pfeil) Reaktion entsteht, ist gleich groß, da ein Enzym die Aktivierungsenergie einer Reaktion lediglich senkt und daher keinen Einfluss auf die freigesetzte Energie ΔG nimmt.

Ihren Namen und ihre spezielle EC-Nummer, erhalten Enzyme normalerweise von der Substanz, die sie umsetzen und von der katalytischen Reaktion, die sie katalysieren. Je nach Reaktionsart können Enzyme einer der folgenden sechs Enzymklassen zugeordnet werden: Oxidoreduktasen, Transferasen, Hydrolasen, Lyasen, Isomerasen und Ligasen. Die Oxidoreduktasen fungieren als Elektronenüberträger und auch als Reduzierer. Daher katalysieren sie die

Redoxreaktionen Oxidation und Reduktion. Ein Vertreter dieser Enzymklasse ist die Lactatdehydrogenase, welche Pyruvat zu Lactat reduziert. Transferasen, wie die Hexokinase, hingegen übertragen komplette funktionelle Gruppen, wie beispielsweise Aminogruppen oder auch Phosphatgruppen. Die Hexokinase kann nur Hexosen phosphorylieren. Werden zum Beispiel Ester- oder Peptidbindungen unter Wasseraufnahme (hydrolytisch) gespalten, so handelt es sich um die Klasse der Hydrolasen. Ein Vertreter der Hydrolasen ist die Triglyceridlipase. Sie spaltet Neutralfette. Lyasen können an bestehenden Doppelbindungen addieren. Zusätzlich sind sie in der Lage Kohlenstoff-Kohlenstoff, Kohlenstoff-Sauerstoff und Kohlenstoff-Stickstoff Bindungen zu spalten, wobei eine Doppelbindung entsteht. Die Glucosephosphat-Isomerase isomerisiert Glucose-6-phosphat zu Fructose-6-phosphat und somit zählt zu der Klasse der Isomerasen. Sie katalysieren strukturelle Veränderungen innerhalb eines Moleküls, welche den Bruch und die Neubildung von kovalenten Bindungen erfordert. Ligasen knüpfen Verbindungen zwischen zwei Substraten, unter Spaltung von Adenosintriphosphat (ATP) [3][4].

Des Weiteren zeigen Enzyme eine Substratspezifität. Darunter ist zu verstehen, dass ein bestimmtes Enzym nicht jede beliebige Reaktion katalysiert, sondern nur bestimmte Substrate zu bestimmten Produkten umsetzen kann. Im Purinkatabolismus beispielsweise existiert eine Abbaureaktion, in welcher das optisch aktive (L)-Allantoin mit Hilfe einer enzymatischen Reaktion zu Allantoat umgewandelt wird. Bei diesem Enzym handelt es sich um Allantoinase, welches genauer gesagt eine Substrat-Stereospezifität aufweist, da es nur das (L)-Allantoin als Substrat verwenden kann.

Damit Enzyme auf die jeweiligen Substrate wirken können, müssen sie es binden. Daher besitzen Enzyme Substratbindungsstellen, die so genannten aktiven Zentren, welche sich in einer Art Tasche befinden. Im aktiven Zentrum stehen die an der Reaktion beteiligten Aminosäuren und auch, falls benötigt, Co-Faktoren (z.B. Metalle) zur Verfügung. Es existieren zwei Mechanismen zur Bindung des Substrates, welche in der nachfolgenden Abbildung dargestellt sind [3].

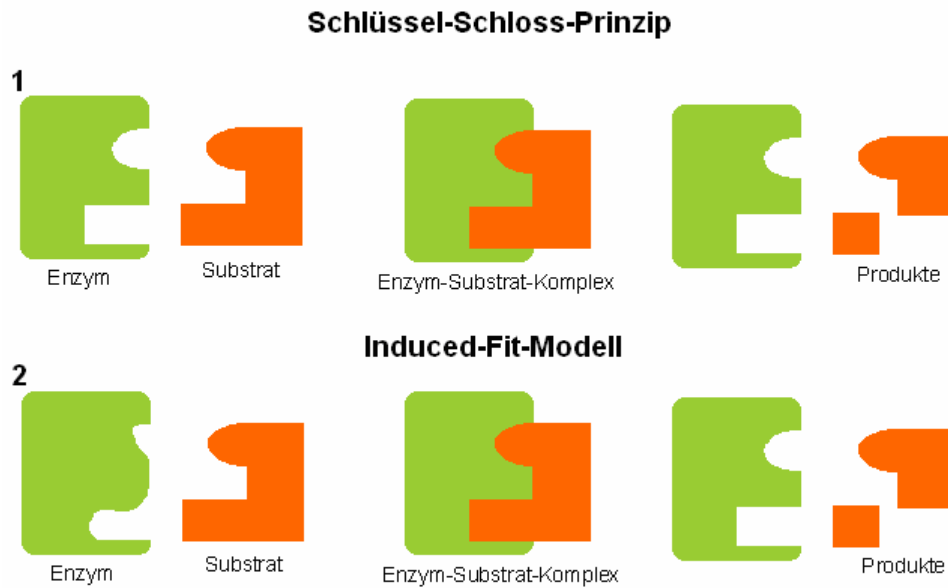


Abbildung 2 1) Schloss-Schlüssel-Prinzip; 2) Induced-Fit-Modell, Substratbindungsstelle nur unvollständig vorgeformt [Quelle: nach Vorlage von [3]]. Grün: Enzym, Orange: Substrat.

Zum einen kann ein Substrat nach dem Schlüssel-Schloss-Prinzip, zum anderen nach dem Induced-Fit-Modell binden. Das Schlüssel-Schloss-Prinzip besagt, dass sich die Raumstruktur der Substratbindungsstelle zu der des Substrates wie ein Schlüssel zum Schloss verhält. Während hingegen die Substratbindungsstelle im Induced-Fit-Prinzip noch unvollständig geformt vorliegt und erst durch die Anwesenheit des Substrates die determinierte Konformation annimmt.

Die Aktivität von Enzymen hängt neben der Konzentration von Substrat und Produkt, der Menge an Enzym und der Temperatur auch vom pH-Wert ab. Die pH-Wert Abhängigkeit beruht auf der Tatsache, dass im aktiven Zentrum vieler Enzyme häufig ionisierbare Aminosäurereste sitzen, deren Ladungszustände sich mit dem pH-Wert ändern. Für die Enzyme gibt es daher pH-Wert Optima. Zum Beispiel liegt das Optimum vom Enzym Pepsin in einem sauren pH-Wert Bereich von 2,0. Trypsin hingegen hat ein basisches Optimum von 8,0. Bei der Temperaturabhängigkeit treten ebenfalls bestimmte Temperaturen auf, in denen die Enzymaktivität am größten ist. Hierbei wird von einem Temperaturoptimum gesprochen [3].

1.1.2 Aufbau und Struktur der Proteine

Proteine bestehen aus 20 universell verwendeten *L*-Aminosäuren. Eine Aminosäure setzt sich aus zwei funktionellen Gruppen und einer für sie speziellen Seitenkette zusammen. Bei den funktionellen Gruppen handelt es sich zum einen um eine Aminogruppe ($-NH_2$) und zum anderen um eine Carboxylgruppe ($-COOH$). Diese beiden Gruppen bilden das so genannte Rückgrat einer Aminosäure. Dieses Rückgrat besitzt ein zentrales C-Atom (C_α -Atom). Zusätzlich zu dem Rückgrat weist eine Aminosäure einen individuellen Rest auf, welcher die für jede Aminosäuren spezielle Seitenkette verkörpert und am C_α -Atom des Rückgrates hängt. Aufgrund der von Aminosäure zu Aminosäure unterschiedlichen Seitenketten, bestimmen diese die physikalisch-chemischen Eigenschaften (Polarität, Hydrophobizität, Säure-Base-Verhalten) einer jeden Aminosäure (siehe Abbildung 3). Die Aminogruppe kann in zwei enantiomeren (stereoisomeren) Formen, der *L*-Form bzw. *D*-Form, vorkommen. *D*-Aminosäuren kommen aber ausschließlich in kurzen Peptiden vor.

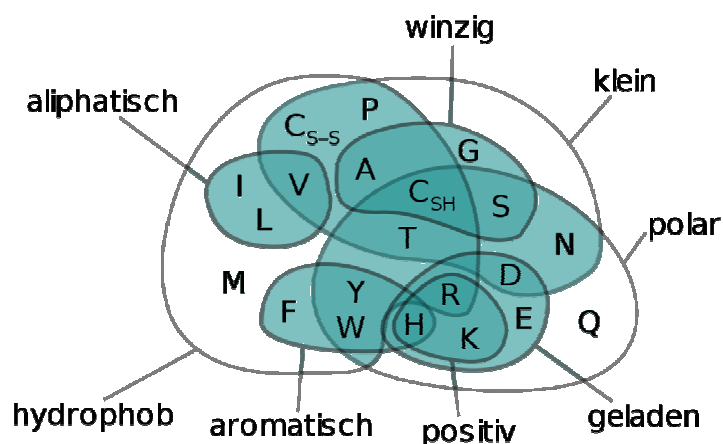


Abbildung 3 Venn Diagramm der Aminosäure Eigenschaften.
Gleiche Eigenschaften werden in Gruppen zusammengefasst, dabei können sich Eigenschaften überschneiden. Bsp.: Aliphatisch = I, L, V. bzw. V = aliphatisch, hydrophob, winzig
[Quelle: <http://de.academic.ru>]

Der Weg von der Verkettung der Aminosäuren bis zu einem vollständigen Protein führt über einzelne Stufen:

Primärstruktur. Durch eine Kondensationsreaktion von zwei Aminosäuren bildet sich eine kovalente Peptidbindung zwischen ihnen. Die Kondensationsreaktion findet zwischen der Carboxylgruppe der Aminosäure 1 und der Aminogruppe der

Aminosäure 2 statt (siehe Abbildung 4). Infolge einer Verlängerung des entstandenen Dipeptids mit weiteren Aminosäuren entsteht eine Polypeptidkette. Diese Polypeptidkette stellt die Primärstruktur eines Proteins dar, in der jede Aminosäure eine eindeutig bestimmte Position besitzt.

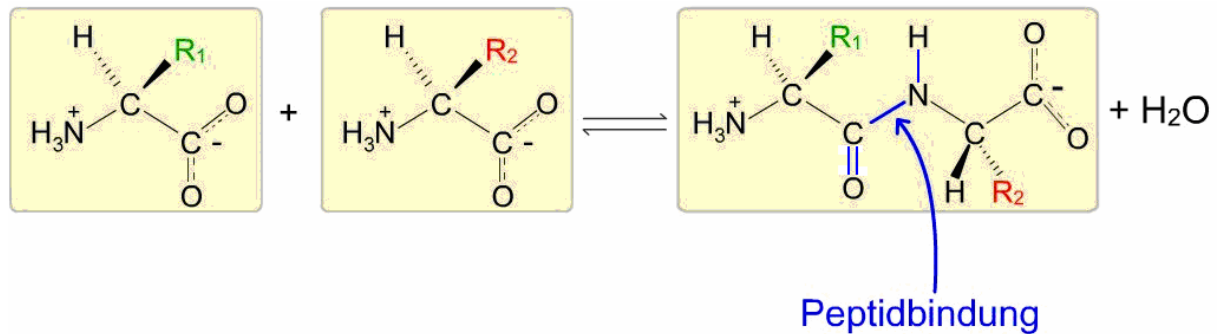


Abbildung 4 Bildung einer **Peptidbindung zwischen zwei Aminosäuren. **R1**, **R2** stellen die Reste der Aminosäure 1 bzw. 2 dar, welche sich in der Kondensationsreaktion nicht verändern. Während der Reaktion wird Wasser (H_2O) abgespalten. [Quelle: <http://www.chemgapedia.de>]**

Die Primärstruktur eines Proteins wird von der DNS codiert. Kommt es in der Primärstruktur zu einer Mutation (Insertion/ Deletion) der enzymatisch notwendigen Aminosäuren, so verändert sich auch die Tertiärstruktur des Proteins. Welches in einer schweren Funktionsstörung resultieren kann. Funktionsstörungen können Krankheiten wie zum Beispiel Unverträglichkeiten und Krebs verursachen. Die Primärstruktur wird üblicherweise als Abfolge des Ein-Buchstaben-Codes oder des Drei-Buchstaben-Codes (siehe Tabelle1) der Aminosäuren dargestellt. Letztendlich legt die Primärstruktur die Sekundär- und Tertiärstruktur fest [3][5][7].

Sekundärstruktur. Die Sekundärstruktur stellt regelmäßige Unterstrukturen des Hauptkettenverlaufs einer Peptidkette dar. Diese Unterstrukturen beruhen auf den physikalisch-chemischen Eigenschaften der Aminosäuren und entstehen durch die Faltung der Peptidkette und der somit hervorgerufenen räumlichen Konformation. Die entstehenden Faltungsmuster sind abhängig von den Eigenschaften der Aminosäuren und der daraus resultierenden intermolekularen Kräfte zwischen den Aminosäuren. Entsprechend werden die Faltungsmuster mit α -Helix, β -Faltblatt, Loop und Turn bezeichnet (Helix, Faltblatt siehe Abbildung 6). Die Entstehung helikaler Strukturen gehen aus den, durch mehrere aufeinander folgende Aminosäuren resultierenden, konstanten Torsionswinkeln Ψ und Φ hervor (siehe Abbildung 5).

Es können sich Proteine falten, welche einzig und allein aus α -Helices bzw. aus β -Faltblättern bestehen. Des Weiteren existieren so genannte Sandwich-Proteine,

diese bestehen aus einem Teil reiner helikaler Strukturen und zum anderen Teil aus reinen Faltblattstrukturen.

Tabelle 1 Tabellarische Übersicht des Dreibuchstabencodes und des Einbuchstabencodes der 20 Aminosäuren. [Quelle: Nach Vorlage von [5]]

Name	Dreibuchstabencode	Einfachbuchstabencode
Alanin	Ala	A
Arginin	Arg	R
Asparagin	Asn	N
Asparginsäure	Asp	D
Cystein	Cys	C
Glutamin	Gln	Q
Glutaminsäure	Glu	E
Glycin	Gly	G
Histidin	His	H
Isoleucin	Ile	I
Leucin	Leu	L
Lysin	Lys	K
Methionin	Met	M
Phenylalanin	Phe	F
Prolin	Pro	P
Serin	Ser	S
Threonin	Thr	T
Tryptophan	Trp	W
Tyrosin	Tyr	Y
Valin	Val	V

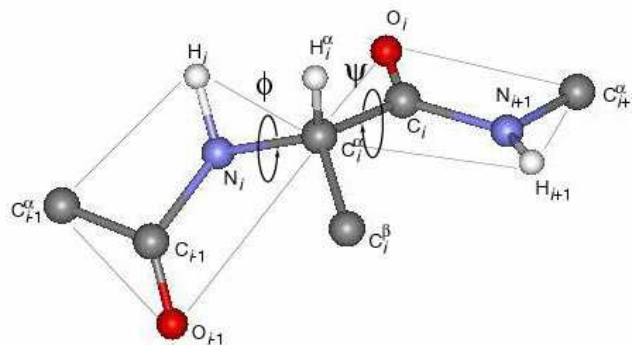


Abbildung 5 Schema der dreidimensionalen Struktur der Peptidbindung mit Drehrichtung und Torsionswinkel (Rotationswinkel) Φ und Ψ . [Quelle: <http://www.chemgapedia.de>]

Helices werden mittels zusätzlichen Wasserstoffbrückenbindungen stabilisiert. Diese bilden sich zwischen einem Wasserstoffatom, welcher an ein Kohlenstoff-, Stickstoff-

oder Sauerstoff gebunden ist und einem gegenüberstehenden Wasserstoffatom, welches dieselben Bedingungen erfüllen muss, aus. β -Faltblätter entstehen, wenn mindestens zwei, zumeist antiparallel verlaufende Stränge miteinander mittelst Wasserstoffbrückenbindungen verbunden sind, welche sich zwischen der Carboxylgruppe der Aminosäure 1 und der gegenüberliegenden Aminogruppe der Aminosäure 2 bildet [3][5].

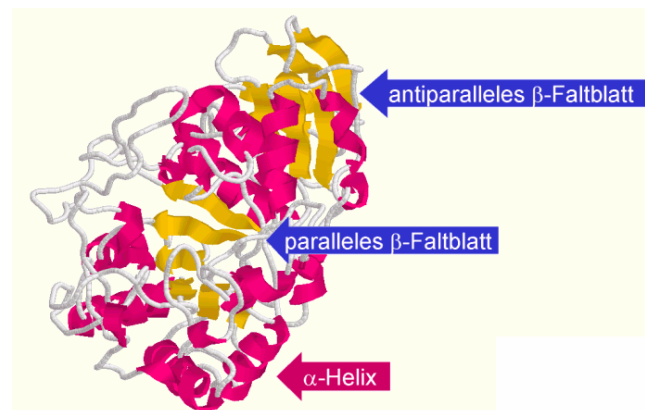


Abbildung 6 3D-Darstellung des im Speichel vorkommenden Enzyms Amylase. Pink = α -Helices, gelb = β -Faltblätter. β -Faltblätter können parallel oder antiparallel zu einander verlaufen. [Bildquelle: <http://www2.chemie.uni-erlangen.de>].

Tertiärstruktur. Die Tertiärstruktur stellt die dreidimensionale Struktur eines Proteins dar. Sie beinhaltet Atome, Aminosäureseitenketten und Sekundärstrukturelemente (z.B. α -Helix), welche nun räumlich zueinander ausgerichtet werden und in Wechselwirkung treten. Stabilisiert wird die Tertiärstruktur eines Proteins durch nicht kovalente Wasserstoffbrückenbindungen, Ionenbeziehungen, hydrophobe Beziehungen und kovalente Disulfidbrückenbindungen sowie der räumlichen Anordnung der Aminosäuren untereinander. Die sehr stabilen Disulfidbrückenbindungen können sich nur zwischen den schwefelhaltigen Aminosäuren Cystein und Methionin ausbilden. Eine Tertiärstruktur stellt somit eine funktionelle Untereinheit (Monomer) dar [3][5].

Quartärstruktur. Wenn sich mehrere Tertiärstrukturen räumlich aneinander anordnen und sich somit mehrere Monomere zu einem Dimer bzw. Oligomer zusammensetzen wird die Quartärstruktur eines Proteins gebildet. Die meisten Proteine werden erst ab dieser Strukturebene funktionsfähig. Die einzelnen Monomere werden mittels Wasserstoffbrückenbindungen und hydrophoben Beziehungen verknüpft [7].

1.1.3 Proteinfaltung

Der Begriff Proteinfaltung bezeichnet den Übergang vom linearen, ungefalteten, nicht-nativen Zustand der Aminosäuren (AS) zu einem nativen, funktionsfähigen Zustand. In diesem Übergang bildet eine Aminosäuresequenz eines Proteins ständig ein und dieselbe typische Tertiärstruktur aus. Im Augenblick sind jedoch die genaueren Mechanismen und Abläufe der Proteinfaltung größtenteils ungewiss.

Levinthal-Paradox. Geschieht der Prozess der Proteinfaltung rein zufällig oder folgt er einem vorgegeben Pfad (Abbildung 7)? Wie kann eine lineare Polypeptidkette in die für sie korrekte gefaltete Konformation (Tertiärstruktur) übergehen? Die unterschiedlichen Konformationen eines Proteins beruhen auf den verschiedenartigen Ausrichtungen der Torsionswinkel (Rotationswinkel) um die Bindung des Peptidrückgrats (siehe Abbildung 5) und der Aminosäure-Seitenketten. Bei einigen Proteinen gibt es zusätzlich die Möglichkeit Disulfidbrückenbindungen auszubilden (nur AS Cystein & Methionin). Es wird davon ausgegangen, dass in einer Polypeptidkette alle Reste j mögliche Konformationen einnehmen können, bei n Peptidbindungen für $n+1$ Aminosäurereste, sodass die Summe aller möglichen Faltungsmuster für das gesamte Polypeptid j^n ist. Basierend auf thermodynamischen Analysen entspricht $j = 8$. Dies zu Grunde liegend besitzt ein kleines Polypeptid, bestehend aus 100 Aminosäuren, 10^{89} mögliche Faltungsmuster. Die durchschnittliche Proteinfaltungszeit t , die ein Polypeptid benötigen würde, um alle vorstellbaren Konformationen eingehen zu können, wenn die Geschwindigkeitskonstante für den Übergang von einer mögliche Konformation zur nächst möglichen k_v ist, ergibt sich aus folgender Gleichung (Levinthal Formel):

$$t = (n \cdot k_v)^{-1} \cdot j^n$$

Formel 1 Levinthal-Formel.

Da die Geschwindigkeitskonstante k_v nicht größer als 10^{-13} s^{-1} betragen kann, würde selbst ein Polypeptid, bestehend aus 100 Aminosäuren, mehr als 10^{66} Jahre benötigen, um alle erdenklichen Faltungsmuster zu erproben. Folglich kann ein Protein nicht zufällig aus dem ungefalteten Zustand in die richtige Konformation übergehen indem es alle denkbaren Faltungen ausprobiert [8].

Bei der Betrachtung von frei im Cytoplasma vorkommenden Aminosäureketten, stellt sich heraus, dass aufgrund der hydrophoben Charakteristika der Aminosäuren,

Bereiche der Proteinsequenz stärkeren Wechselwirkungen mit der wasserähnlichen Umgebung existieren. Diese Bereiche weisen höhere Energien als andere Abschnitte der Sequenz auf. Ein Protein im nicht-nativen Zustand besitzt also ein höheres Maß an freier Energie, als eines im nativen Zustand. Aufgrund des Prinzips der Thermodynamik basierend auf der Entropie (Ordnungszustand) wird im nicht-nativen Protein stets der Zustand der geringsten freien Energie eingenommen. Folglich wird das Protein die Konformation mit der geringsten freien Energie einnehmen. Es wird angenommen, dass sich zunächst kleine Sequenzabschnitte falten und so genannte Faltungsnuclei bilden. Diese nehmen ihre dreidimensionale Faltung ein, um welche sich anschließend die restlichen Strukturbereiche winden. Wichtig für die Betrachtung der Proteinfaltung sind, wie oben erwähnt, die hydrophoben Eigenschaften der Aminosäuren, denn die Lösemittel, welche globuläre Proteine umgeben, besitzen meist einen hohen Wasser Anteil. Dadurch befinden sich die hydrophilen Reste meist im Proteininneren, während sich die hydrophoben Reste auf der äußeren Oberfläche des Proteins aufhalten. Daher ist es von großer Bedeutung eine Bildungsvorschrift für die Faltung von Proteine zu entwickeln, die die energetischen Zustände einer Aminosäure innerhalb der Peptidkette und die daraus resultierenden Aminosäure-Aminosäure- bzw. Aminosäure-Lösemittel-Wechselwirkungen mit einzubeziehen [5].

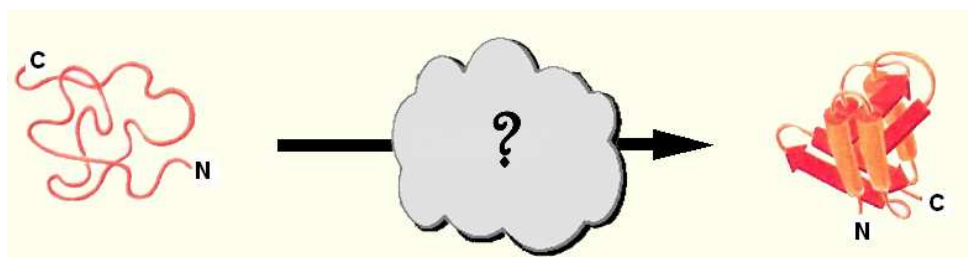


Abbildung 7 Faltungsweg eines Polypeptides. Prozess beginnt im ungefalteten, kettenförmigen Zustand der Aminosäuresequenz und führt über einen bestimmten Faltungscode (Blase mit Fragezeichen) zu einem funktionsfähigen Protein im nativen Zustand [Quelle: nach Vorlage von [8]].

1.2 Bioinformatische Grundlagen und Werkzeuge

1.2.1 Protein Familien Datenbank

Die ProteinFamilies Datenbank (Pfam Database) ist eine umfangreiche Sammlung bzw. Zusammenstellung von Protein-Domänen und Protein-Familien, welche von Multiplen Sequenzalignments und Hidden Markov Model-Profilen (HMM Profiles) repräsentiert werden [9]. HMMs, erstellt mit Hilfe von HMMER, beschreiben, im Gegensatz zu positionsspezifischen Matrizen (PSSM), die Wahrscheinlichkeiten für alle Übergangszustände an jeder Position in der Sequenz. In einer PSSM hingegen wird eine Konsensussequenz abgebildet, an hand derer man für jede Position in der Sequenz die Wahrscheinlichkeit erhält, mit der dort jede Aminosäure vertreten ist [11].

The screenshot displays the Pfam website interface for the OHCU_decarbox (PF09349) family. The top navigation bar includes links for HOME, SEARCH, BROWSE, FTP, HELP, and ABOUT. The Pfam logo is in the top right corner. Below the navigation bar, the family name 'Family: OHCU_decarbox (PF09349)' is prominently displayed. To the right of the family name, statistics are shown: 12 architectures, 433 sequences, 1 interaction, 241 species, and 19 structures. A left sidebar menu lists various options: Summary (selected), Domain organisation, Alignments, HMM logo, Trees, Curation & models, Species, Interactions, and Structures. Below this menu is a 'Jump to...' section with an 'enter ID/acc' field and a 'Go' button. The main content area is titled 'Summary' and contains a description of the OHCU decarboxylase family, literature references, an InterPro entry (IPR018020), and external database links (PANDIT, Pseudofam, SYSTERS). On the right side of the summary, there is a 3D molecular structure visualization of the OHCU decarboxylase in complex with guanine, labeled as 'Example structure' and 'PDB entry 2o74'.

Abbildung 8 Die Pfam Familien Seite der Webseite. Das Bild zeigt die *Summary* Information der OHCU Decarboxylase Familie (PF09349). Im linken Menü kann der Besucher zwischen Zusatzinformations- Bereichen wählen, zB. Trees oder Species. „Jump to...“ kann benutzt werden, um auf eine andere Familie oder einen anderen Eintrag zugelingen. Das Panel-Menü oben rechts gibt eine Zusammenfassung über die Anzahl der in der Familien befindlichen Architekturen, Sequenzen, Interaktionen, Spezies und Strukturen. Das rechts dargestellte 3-D-Bild zeigt eine Struktur der OHCU Decarboxylase. [Quelle: <http://pfam.sanger.ac.uk>]

Des Weiteren enthält die Pfam Datenbank wichtige Informationen zu Architektur, Interaktionen sowie Strukturen (siehe Abbildung 8). Die auch als Motiv-Datenbank geltende Pfam-Datenbank, wird in Pfam A und Pfam B unterteilt. Jede Pfam A

Familie besteht aus einem *Seed* Alignment, welches ein kleines Set an repräsentativen Mitgliedern der Familie bietet, auf den *Seed* Alignments basierende Hidden Markov Modelle (HMM, siehe Abbildung 9) und einem automatisch erstellten *Full* Alignment, welches alle zur Familie gehörenden und detektierbaren Proteinsequenzen enthält. Das größte *Full* Alignment enthält über 68000 Sequenzen. Weiter werden Pfam A Familien mit einer permanenten Zugangsnummer (das Kürzel PF gefolgt von einer fünfstelligen Zahl: PFxxxxx), einem ID Namen und einer Definitionszeile versehen. Pfam B Einträge werden automatisch aus der PRODOM-Datenbank generiert und als Sequenz - Familien - *Single* Alignments dargestellt. Zusätzlich werden alle Pfam B Einträge mit dem Kürzel PD und einer fünfstelligen Zahl benannt (PDxxxxx). Die Qualität der Pfam B Alignments ist nicht ausreichend genug um brauchbare HMMs zu erzeugen, daher erfüllt Pfam B eher den Zweck einer Ablage bzw. eines Speichers von homologen Informationen und einer Reserve für noch uncharakterisierte Proteinfamilien. Das Ziel der Pfam Datenbank ist es nun die größten Familien aus Pfam B progressiv in Pfam A einzugliedern [10]. Alle Pfam-Daten sind über das Internet frei zugänglich und als Download in verschiedenen Formen erhältlich [9].

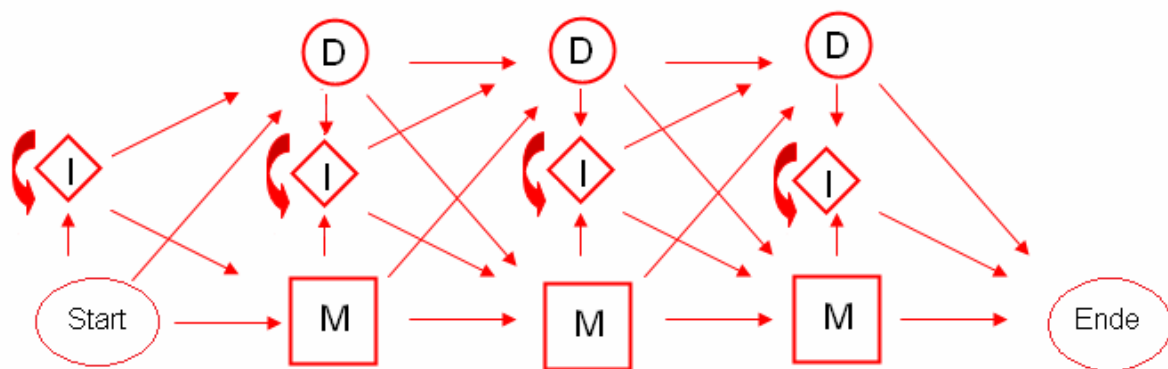


Abbildung 9 Beispiel eines Hidden Markov Modells (HMM) für mögliche Zustände in einem multiplen Alignment. M: Match (Übereinstimmung), I: Insertion (Einfügung), D: Deletion (Löschung) [Quelle: nach Vorlage von [11]].

Die anfänglichen Mitglieder eines *Seed* Alignment werden aus verschiedenen Quellen (z.B. Swissprot, Prosite, PRODOM) bezogen. Enthält eine der Quellen ein vollständiges Alignment dieser Seed-Mitglieder so wird dies benutzt, andernfalls wird ein Alignment, indem bekannte und auffällige Charakteristika, wie am aktiven Zentrum beteiligte Aminosäuren oder strukturell wichtige Reste, gebaut. Die

Alignment Methode Clustal W (siehe 1.2.2) liefert am häufigsten das beste Alignment. Das erzeugte Seed Alignment wird nun verwendet um HMMs zu erstellen und damit neue Mitglieder aufzuspüren. Anschließend wird ein Alignment aller detektierten Mitglieder generiert. Wenn einmal das *Seed* und das *Full* Alignment einer Familie konstruiert wurden, wird eine Reihe von Qualitätskontrollen durchgeführt. Es werden Falsch-Positive und Falsch-Negative zu einem Referenz-Cluster, für gewöhnlich aus PROSITE stammend, überprüft. In dieser Überprüfung wird sicher gestellt, dass sich weder das *Seed* noch das *Full* Alignment in irgendeinem Rest mit einer anderen Familie überschneidet. Der Vorteil der *Seed* und *Full* Alignments besteht darin, dass es die Stabilität der Datenbank sichert. Denn das manuelle *Seed* Alignment entspricht einer hohen Qualität und die *Full* Alignments resultieren aus den *Seed* Alignments. Produziert ein Seed Alignment kein nützliches HMM so wird es solange verbessert bis ein befriedigendes Ergebnis erreicht wird [10].

Seit 2005 existierten Pfam-Clans. Ein Clan ist eine Zusammenstellung von Pfam A Einträgen, welche als homolog beurteilt werden. Die Pfam-Clans werden manuell erstellt und basieren auf verschiedensten Informationsquellen, wie beispielsweise Primärstrukturen, bekannten Strukturen, Profil-Profil-Vergleichen und anderen Datenbanken (z.B. SCOP). Es wird versucht die Beziehungen zwischen den Mitgliedern der einzelnen Familien graphisch mittels Clan-Alignments darzustellen. Clans definieren folglich eine einfache hierarische Klassifikation von Pfam Einträgen und erlauben einen besseren Transfer von strukturellen bzw. funktionellen Informationen zwischen Familien. Demzufolge ist eine bessere Vorhersage von Funktion und Struktur für Familien mit unbekannter Funktion möglich. Um die Beziehungen zwischen den Familien abzuleiten werden drei verschiedene Tools benutzt: Profil-Profil- Vergleichs-Tool PRC, Profil-Profil-Vergleichs-Tool HHsearch und ein einfacher Vergleich eines Output Programms SCOOP. Die Kombination dieser drei Tools erlaubt ein sensitives Untersuchen, da jedes Tool sensitiv für ein anderes Set an Beziehungen ist, und die Kombination verstärkt zugleich unabhängig voneinander detektierte Beziehungen [9].

Die Pfam 24.0 Datenbank besitzt gegenwärtig mehr als 11900 Familien [12].

1.2.2 Multiples Sequenz Alignment

Das multiple Alignment bezeichnet das gleichzeitige Analysieren mehrerer Sequenzen und lehnt sich an das paarweise Alignment an. Durch den Vergleich von zwei und mehr Sequenzen werden Unterschiede und Ähnlichkeiten deutlich, welche beispielsweise Rückschlüsse auf funktionelle, strukturelle und evolutionäre Beziehungen zeigen. Denn aufgrund des evolutionären Drucks werden Veränderungen bzw. Mutationen in den Aminosäuresequenzen verursacht. Daher unterliegen Proteine einem ständigen Funktions- und Strukturwandel. Das multiple Alignment liefert wichtige Informationen über Aminosäureverteilungen an einzelnen Positionen innerhalb der Sequenzen. Derartige Verteilungen geben Aufschluss über Konservierungen einzelner Aminosäuren bzw. gesamter Bereiche. Des Weiteren bilden diese Aminosäureverteilungen die Grundlage für profilbasierte Datenbanksuchen und phylogenetische Analysen. Um Multiple Alignments zu erstellen existieren globale und lokale Berechnungs-Methoden. In einem globalen Alignment werden die Sequenzen, unter Einführung von Lücken (Gaps), zu einem Block zusammengefasst. Das lokale Alignment hingegen sucht in den Sequenzen nach Bereichen großer Ähnlichkeit und erstellt dann blockweise multiple Alignments ohne Verwendung von Lücken [11].

Eine Methode zur Berechnung eines globalen Alignments lieferten 1987 Feng & Doolittle, welche den Needleman & Wunsch Algorithmus zu Grunde zieht, und eine weitere Methode wurde 1994 von J. Thompson, D. Higgins & T. Gibson entwickelt – Clustal W [11].

Clustal W. Der Clustal W Algorithmus berechnet ein progressives und globales multiples Alignment. Aufgrund der differenzierten Bewertung von Lücken und Aminosäuren wird diese Methode sensitiver als die von Feng & Doolittle [11].

Der Clustal W Algorithmus lässt sich in drei Schritte untergliedern. Im ersten Schritt werden die Distanzen berechnet. Hierbei wird zunächst eine Distanzmatrix (siehe Abbildung10) erstellt, dies geschieht entweder *slow* oder *fast*. Der schnelle heuristische Weg (*fast*) geht über die Anzahl der übereinstimmenden Treffer bei einer *k-tuple-Suche*. In der Mathematik bezeichnet ein *Tuple* ein Element in einer geordneten Menge. Wir also k der Wert 3 zugeteilt, so sucht der Algorithmus in der Matrix nach einem exaktem Match der Länge 3. Diese *k-tuple-Suche* liegt auch der im Dotplot verwendeten Wort-Filter-Methode zu Grunde. Das *slow* Verfahren erstellt dagegen ein optimales paarweises Alignment nach Needleman & Wunsch. Das

heißt, dass zunächst eine zweidimensionale Matrix (ähnlich einem Dotplot) mit den zu vergleichenden Sequenzen erstellt wird. Dadurch kann nun jedes Alignment mittels eines Pfades durch die Matrix beschrieben werden. Das beste Alignment ist das mit der höchsten Endsumme. Die Endsumme (Score) ergibt sich aus der Addition der Werte in jeder einzelnen Matrixzelle. Der jeweilige Einzelwert der Matrixzelle stammt hierbei aus der verwendeten Substitutionsmatrix, welche im einfachsten Fall nur von der Identität abhängt. Dabei wird die Identität beispielsweise mit 1 bewertet, andernfalls mit 0. Des Weiteren wird die Bildung einer Lücke mit in die letztendliche Score-Bildung einbezogen. Denn sie entsteht immer dann, wenn der diagonale Weg in der Matrix verlassen wird. Dies führt zu einer Verlängerung des Weges und zum Einführen von Lücken, welche jedes Mal Strafpunkte kosten. Clustal W benutzt mehrere Substitutionsmatrizen in direkter Abhängigkeit von der Distanz zwischen den Aminosäuren. Folglich werden eng verwandte Sequenzen mit einer BLOSUM80 und eher weit voneinander entfernte Sequenzen mit einer BLOSUM30 bewertet. [11].

Seq 1	0						
Seq 2	0,17	0,00					
Seq 3	0,59	0,60	0,00				
Seq 4	0,59	0,59	0,13	0,00			
Seq 5	0,77	0,77	0,75	0,75	0,00		
Seq 6	0,81	0,82	0,73	0,74	0,80	0,00	
Seq 7	0,87	0,86	0,86	0,88	0,93	0,90	0,00
	Seq 1	Seq 2	Seq 3	Seq 4	Seq 5	Seq 6	Seq 7

Abbildung 10 ClustalW. Distanzmatrix von 7 Sequenzen – Berechnete Distanzen der Seq 1 bis Seq 7 zueinander [Quelle: nach Vorlage von [11]].

Im zweiten Schritt des Clustal W Algorithmus wird ein Neighbor-Joining Initialbaum berechnet. Hierbei wird der Baum ausgehend von der Distanzmatrix (siehe Abbildung10) in zwei Stufen berechnet. Von dem Initialbaum ausgehend werden anschließend die Sequenzen untereinander verglichen. Die Distanzen werden mit der Neighbor-Joining Methode (NJ) gebildet. Hier werden die Sequenzen zunächst an den Ästen des ersten Baumes mittels Sterntopologie verteilt (siehe Abbildung 11). Dabei spiegeln die einzelnen Astlängen die Distanzen der jeweiligen Sequenzen wieder. Die Mitte des Baumes stellt die Wurzel dar. Von dieser ausgehend wird der gewurzelte NJ-Baum konstruiert, bei dem jeder Ast mit einem geeigneten Verfahren speziell gewichtet wird. Dabei erhalten Gruppen mit eng verwandten Sequenzen

einen kleinen Wert und einzeln stehende Sequenzen mit einer großen Distanz zu allen anderen einen hohen Wert.

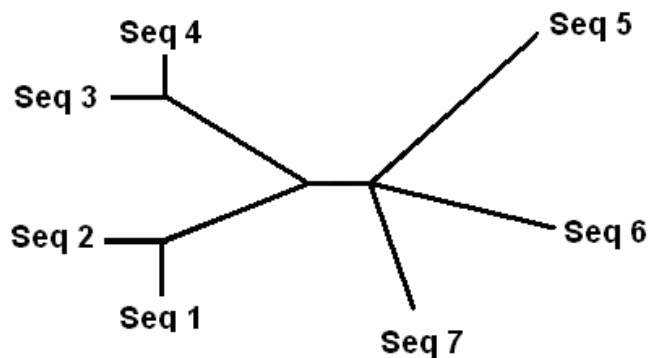


Abbildung 11 ClustalW: Ungewurzelter NJ-Baum.
[Quelle: nach Vorlage von [11]]

Wie aus Abbildung 12 erkennbar erhält demgemäß der Ast von Sequenz 7 mit 0,442 die höchste Bewertung, da die Sequenz den größten Abstand zu den anderen Sequenzen aufweist. Die anderen Astlängen ergeben sich aus der Summe der einzelnen Astlängen von der Wurzel zur jeweiligen Sequenz. Hierbei wird jede Länge zuvor durch die Anzahl der an dieser Wurzel befindlichen Äste dividiert. So ergibt sich aus $0,062 : 6 + 0,015 : 5 + 0,0398 = 0,411$ eine Distanz von 0,411 für die Seq 5.

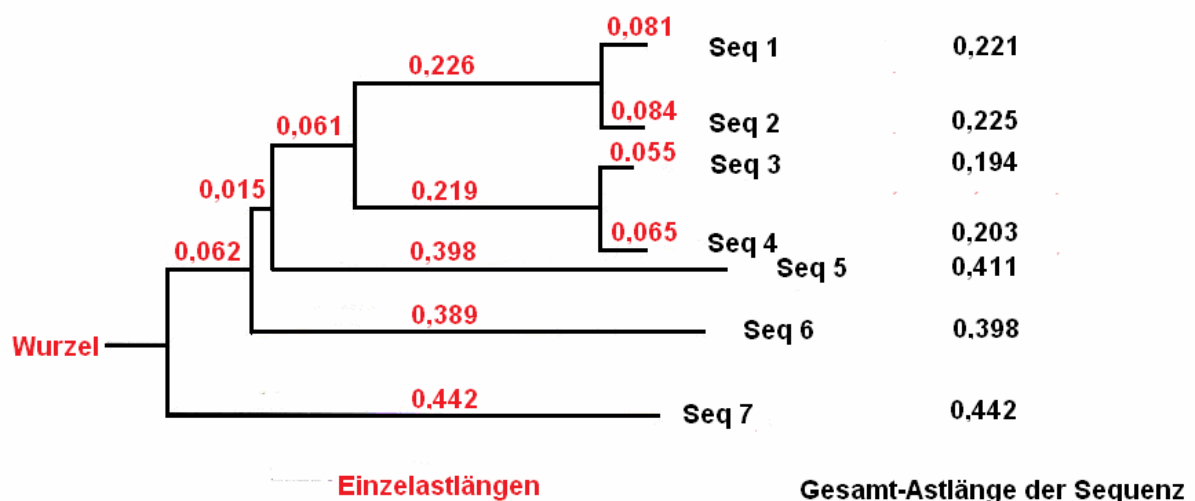


Abbildung 12 ClustalW: Gewurzelter NJ-Baum. Die roten Zahlen entsprechen den Einzelastlängen, die schwarzen am Ende stehenden Zahlen entsprechen der Gesamtastlänge [Quelle: nach Vorlage von [11]].

Im dritten und letzten Schritt wird ein progressives multiples Alignment erstellt. Es beginnt mit den am engsten verwandten Sequenzen. Dies wären in der Abbildung 12 Sequenzen 1 und 2, sowie Sequenzen 3 und 4. Anschließend werden die paarweisen Alignments Sequenz 1/2 und Sequenz 3/4 erneut paarweise verglichen. Darauf folgt dann das Alignment von Sequenz 1/2/3/4 mit Sequenz 5. Bis letztendlich alle Sequenzen align sind. Wird einmal eine Lücke eingeführt, so bleibt diese für immer bestehen. Die Gap-Bewertung richtet sich hierbei nach dem Grad der Verwandtschaft. So ist der Strafpunkt bei Einführung einer Lücke beim paarweisen Vergleich von Sequenzen hoher Identität größer, als bei nicht so eng verwandten Sequenzen. Hinzukommt, dass längere Sequenzen höhere Strafpunkte für Lücken erhalten als kürzere. Wird eine Lücke ausgedehnt, so ist der Wert (Gap-Extension Strafpunkt) von der Differenz der Sequenzlängen zueinander abhängig. Die so ermittelten Strafpunkte werden zusätzlich mit einem empirisch ermittelten und aminosäurespezifischen Faktor multipliziert, welcher sich nach der beobachtbaren Häufigkeit, mit der neben einer Aminosäure Lücken auftauchen, vorkommen [11]. Eine Darstellungsform von Clustal W zeigt die nachfolgende Abbildung: JalView.

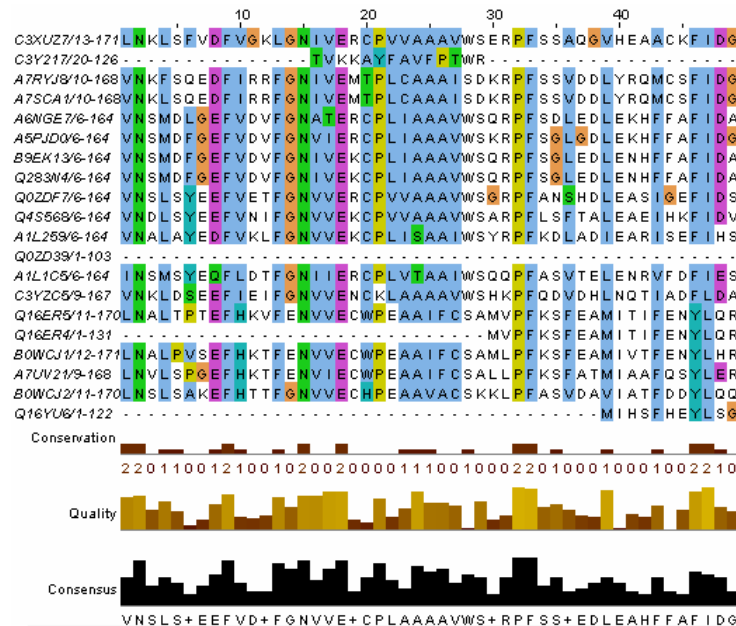


Abbildung 13 Von Clustal W berechnetes MSA, Darstellung in JalView – mit Angabe der Conservation (braun, Konservierung), Quality (hellbraun, Qualität) und Consensus-Sequenz (schwarz), die MSA-Farbgebung entspricht Clustal (siehe 4.1, Abbildung 28) [23].

Wie der Abbildung 13 zu entnehmen, zeigt die MSA Darstellung in JalView neben dem Alignment drei weitere Aspekte: Conservation, Quality und Consensus.

Die Konservierung (Conservation) stellt eine automatisch kalkulierte quantitative Alignment-Annotation dar. Sie misst die Anzahl der konservierten physikochemischen Eigenschaften für jede Spalte des Alignments. Die Angaben über den Grad der Konservierung (Conservation) werden von 0 bis * präsentiert, wobei 0 einer Konservierung von 0% und * einer 100%-igen Konservierung entsprechen. Die quantitative Alignment Qualität (Quality) wird ebenfalls automatisch berechnet. Es handelt sich um eine *ad-hoc* Messung der beobachtbaren Mutationen in jeder Spalte des Alignments. Ein hoher *Alignment Quality-Score* für eine Spalte suggeriert, dass keine Mutation vorkommt oder die meisten Mutationen als favorisiert gelten. Die Consensus-Annotation spiegelt die prozentuale Angabe einer pro Spalte am häufigsten vorkommenden Aminosäure wieder. Betrifft der gleiche Wert zwei oder mehr Aminosäuren an einer Position, so wird dies in der Consensus Zeile mit einem „+“ dargestellt, da stets nur Platz für eine Aminosäure gegeben ist.

1.2.3 Phylogenetische Analyse

Mithilfe phylogenetischer Analysen werden die evolutionären Beziehungen zwischen Organismen versucht aufzuklären. Dies untersuchte früher schon Charles Darwin. Er stützte sich dabei auf morphologische Charakteristika der Organismen. Heute jedoch nimmt die Wichtigkeit der Analysen von Protein- und Nukleotidsequenzen immer mehr zu. Denn die Sequenzanalyse verkörpert die Aufklärung der stammesgeschichtlichen Beziehungen zwischen Lebewesen. In Genen kommt es im Verlauf der Nachfolge-Generationen zu zufälligen Mutationen, meistens Nukleotidsubstitutionen. 1872 schrieb Charles Darwin: „[...] variations, [...] if they be in any degree profitable to the individuals of a species, [...] will tend to the preservation of such individuals, and will generally be inherited by the offspring [...] I have called this principle, by which each slight variation, if useful, is preserved, by the term *Natural Selection*.” Aufgrund der natürlichen Selektion werden nur die Mutationen weitergegeben, die beispielsweise die Funktion eines Proteins nicht negativ beeinflussen. Andernfalls würde die natürliche Selektion Mutationen beseitigen, die die Proteinfunktion stören würden [11][14].

Die drei wichtigsten bioinformatischen Methoden zur Sequenzanalyse bzw. zur Rekonstruktion von evolutionären Bäumen stellen die Distanzmethoden, Maximum-Likelihood-Methoden und die Parsimonymethoden dar. Hierbei existieren, häufig unausgesprochene, aber vorausgesetzte Bedingungen, wie beispielsweise die Annahme der Homologie, d. h. der Ursprung aus einem gemeinsamen Vorfahren, aller Sequenzen eines zu untersuchenden Datensatzes sowie die gemeinsame evolutionäre Entwicklung aller einem Datensatz zugehörigen Sequenzen [11][13].

Topologie phylogenetischer Bäume. Ein phylogenetischer Baum besteht aus Knoten (nodes) und Ästen (branch). Wobei die Knoten in terminal und intern unterschieden werden. Die terminalen Knoten befinden sich am Ende eines Astes und stehen für die untersuchten Gene oder ähnliches, welche meist *operational taxonomic unit* (OPU) genannt werden. Die internen Knoten hingegen, auch als *hypothetical taxonomic unit* (HTU) bezeichnet, vertreten die bereits ausgestorbenen Gene bzw. Spezies (siehe Abbildung 14). Ein phylogenetischer Baum kann als gewurzelter (rooted) oder ungewurzelter (unrooted) Baum dargestellt werden. Das *Rooting* des Stammbaumes, d.h. die Bestimmung einer *Outgroup* (Außengruppe), der

phylogenetisch älteste Ast, bestimmt die Richtung des Stammbaumes und verleiht ihm somit eine historische Bedeutung.

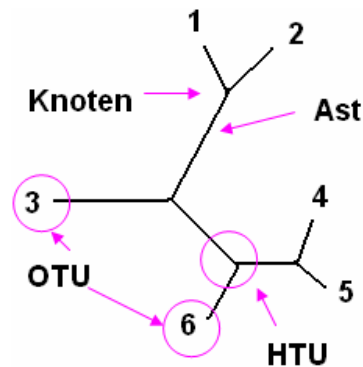


Abbildung 14 Phylogenetischer Baum mit sechs OTUs und vier HTUs [Quelle: nach Vorlage von [11]].

Werden phylogenetische Bäume auf Grundlage von Sequenzen modelliert, so handelt es sich um abgeleitete Bäume, denn dieser Baum wurde mithilfe einer bestimmten Methode basierend auf einem bestimmten Datensatz berechnet [11].

Weiterhin lassen sich die Stammbäume in Kladogramme und Phylogramme unterteilen. Kladogramme zeigen im Gegensatz zu Phylogrammen keine Divergenzzeit der einzelnen Taxa. Die Divergenzzeit wird in Phylogrammen durch die unterschiedlichen Astlängen (siehe Abbildung 15) verdeutlicht.

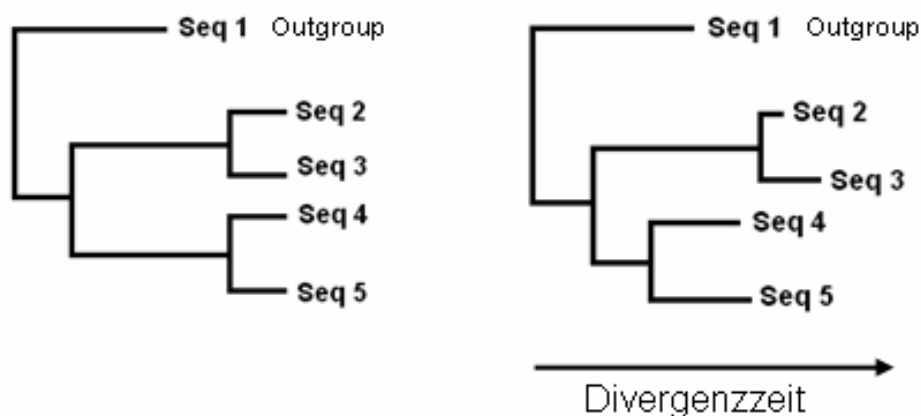


Abbildung 15 Stammbäume. Rechts: Kladogramm; Links: Phylogramm. Seq 1 steht hier für die Outgroup (Außengruppe). [nach Vorlage von Quelle [4]]

In Abbildung 15 kann man die einzelnen Sequenzen in spezielle Gruppen einteilen. Die Sequenzen 2 und 3 sowie die Sequenzen 4 und 5 verhalten sich zu einander monophyletisch. Eine Monophyletische Gruppe umfasst alle Nachfahren einer Spezies. Die Sequenzen 3, 4, und 5 verhalten sich zu einander paraphyletisch, d.h. sie definieren sich über einen gemeinsamen Vorfahr, doch umfasst diese Gruppe nicht alle Nachkommen. Die Sequenzen 3 und 4 gehen auf mehr als einen Vorfahren zurück und werden somit als polyphyletisch bezeichnet.

Stammbaumrekonstruktion. Wie bereits erwähnt, stellen die Distanzmethoden und die Merkmalsmethoden Parsimony- und Maximum-Likelihood die wichtigsten Methoden zur Stammbaumrekonstruktion dar. Alle diese Methoden versuchen einen Stammbaum zu erstellen, der die Entstehung der Taxa im Verlauf der Evolution so genau wie möglich widerspiegelt. Näher betrachtet stellt der entstandene evolutionäre Baum eine Hypothese dar [14]. Dem Ausgangspunkt aller evolutionären Bäume liegt ein multiples Sequenzalignment zugrunde [11].

Im Folgenden wird die Distanzmethode zur Bildung von evolutionären Bäumen genauer beschrieben. Sie bildet die Grundlage für die in Punkt 5.1 dargestellten evolutionäre Bäume.

Die Berechnung von Distanzbäumen geschieht in zwei Schritten. Zunächst wird aus dem multiplen Sequenzalignment eine Distanzmatrix gebildet. Im zweiten Schritt wird der evolutionäre Baum mit UPGMA (Unweight Pair-Group Method Using Arithmetic Averages) bzw. NJ (Neighbor-Joining) kalkuliert. Die UPGMA-Methode setzt eine konstante Evolutionsrate (gleichmäßig und mit konstanter Geschwindigkeit) voraus und benutzt den Mittelwert der Distanzen zur Festlegung der Astlängen. Die NJ-Methode hingegen versucht die Astlängen exakter zu berechnen und einen Baum mit der kürzesten Gesamtlänge zu konstruieren [11].

Erster Schritt. Die Distanz kann mit diversen Methoden berechnet werden. Dazu zählen z.B. die Kimuras Distanz, das Kategorien-Modell und die PAM-Matrix. In der Kimuras Distanz bildet Anzahl der nicht-identischen Aminosäuren zwischen zwei Sequenzen das Distanzmaß. Da dies unabhängig von der Substitution der Aminosäuren geschieht, gilt diese Distanzmethode als eher ungenau, wobei die Bäume nur zur ersten groben Betrachtung herangezogen werden können. Das Kategorien-Modell teilt Aminosäuren in unterschiedliche Kategorien ein. Dieses Modell bewertet den Austausch von einem Purin durch ein anderes Purin bzw. den

Austausch eines Pyrimidins mit einem anderen Pyrimidin (Transition) und den Austausch eines Purins durch ein Pyrimidin bzw. einen Austausch von einem Pyrimidin mit einem Purin (Transversion) unterschiedlich. Somit werden hier nur Aminosäuresubstitutionen betrachtet, welche zu einem Kategorienwechsel führen. Die PAM-Matrix-Methode, basierend auf der PAM1-Matrix, berücksichtigt wie häufig eine Aminosäure und vor allem gegen welche anderen Aminosäuren sie ausgetauscht wurde. Dabei werden konservative Substitutionen anders bewertet als nicht-konservative. Die PAM1-Matrix basiert auf einem globalen Alignment von Proteinsequenzen, welche zu mehr als 85% identisch waren [11].

Zweiter Schritt. Sind nun die Distanzen berechnet, wird aus dieser Distanzmatrix der Baum konstruiert. Dabei steht UPGMA oder NJ zur Wahl.

NJ. Die NJ-Methode wurde schon in Punkt 1.2.2 mit der unterstützenden Abbildung 8 erläutert. Ergänzend hierzu der zugrunde liegende NJ-Algorithmus auf Seite 23.

UPGMA. Die UPGMA-Methode ist ein so genanntes Clusterverfahren, entwickelt von Sokal & Michener 1985. Ausgehend von in der in Schritt 1 berechneten Distanz-Matrix, wird nun ein Dendrogramm erstellt (siehe Abbildung 16). Ähnliche Sequenzen stehen in dem Dendrogramm als Cluster zusammen. Es werden nun zwei Cluster zusammengefasst und das arithmetische Mittel derjenigen Distanzen berechnet. Anschließend wird die nächst ähnlichste Sequenz herangezogen, usw. [11].

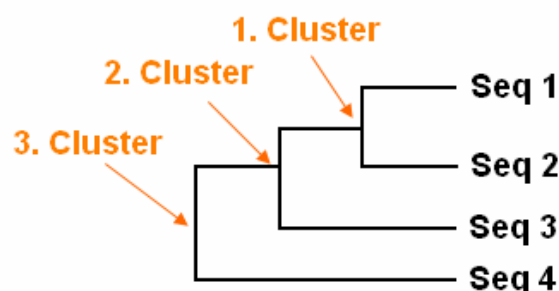


Abbildung 16 Dargestellt ist ein Dendrogramm nach UPGMA.
[Quelle: nach Vorlage von [11]]

Der längste Ast eines mittels UPGMA erzeugten Baumes wird automatisch als Outgroup festgelegt, da UPGMA eine konstante Evolutionsrate voraussetzt [15]. Die Outgroup würde in Abbildung 16 die Seq 4 darstellen.

1 Initialisiere Distanzmatrix $D[1 \dots n, 1 \dots n]$, $E := \{ \}$, $V := L := \{1, \dots, n\}$

Erläuterung: $E \dots$ Kanten $D \dots$ Distanzmatrix [Anzahl der Felder vertikal x horizontal]
 $V \dots$ Vertices $L \dots$ Menge momentan verwendeter Knoten

2 Führe aus.

3 Für jedes Paar $i \neq j \in L$ berechne $d_{ij} := d_j - (r_i \div r_j)$.

Hierbei ist $r_i := \frac{1}{|L|-2} \sum_{m \in L} d_{im}$.

Erläuterung: Für das Paar i, j ($i, j = AS$) wird die Distanz d_{ij}^* berechnet; r_i, r_j entspricht dem mittleren Abstand der anderen $|L|-2$ Einträge zu i und j .

4 Wähle die Indizes i, j ($i \neq j$) so, dass d_{ij}^* minimal ist.

Erläuterung: die zwei Knoten mit dem kleinsten Abstand d zwischen der AS(i) und AS(j) werden identifiziert und zu einem neuen Knoten k zusammengefasst.

5 Definiere neuen Knoten k , füge diesen in V und L ein und lösche i, j aus L

6 Berechne die Abstände $d_{km} := \frac{1}{2}(d_{im} + d_{jm} - d_{ij})$ ($m \in L$) zum neuen Knoten k .

Erläuterung: Für all noch in L verbliebenen Knoten wird die Distanz zu k berechnet

7 Füge die Kanten (i, k) , (j, k) zu E hinzu, die Längen sind

$$d_{ik} := \frac{1}{2}(d_{ij} + r_i - r_j) \text{ und } d_{jk} := d_{ij} - d_{ik}.$$

8 bis L nur noch aus zwei Einträgen i und j besteht. Füge die Kante (i, j) mit Gewicht d_{ij} zu E hinzu.

9 Ausgabe: (V, E)

2. Wechselwirkungen und Energiefunktion

2.1 Intermolekulare Wechselwirkungen

Die physikalisch-chemischen Eigenschaften der Aminosäuren werden durch ihre Seitenketten (Reste) bestimmt. Wie aus dem Venn-Diagramm (siehe Abbildung 3) entnommen werden kann, lassen sich die Aminosäuren anhand ihres Restes klassifizieren und einteilen. Neben Wasserstoffbrückenbindungen, können (bestimmte) Aminosäuren zueinander Disulfidbrücken, Ionenbeziehungen und hydrophobe Bindungen ausbilden, wie die nachstehende Grafik veranschaulichen soll.

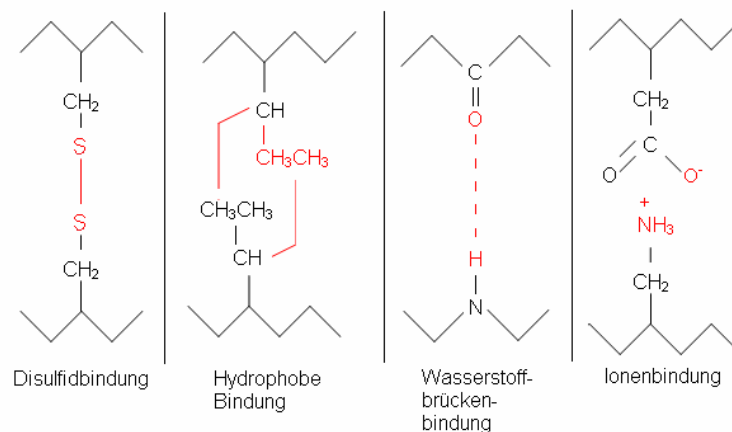


Abbildung 17 Möglichkeiten der Ausbildung von Haupt- und Nebervalenzbindungen bei Proteinmolekülen [Quelle: nach Vorlage von [1]].

In Punkt 1.1.2 wurde bereits das Venn-Diagramm eingeführt. Aus dem Venn-Diagramm lassen sich erste Rückschlüsse auf physikalisch-chemische Eigenschaften zwischen Aminosäuren ableiten. So können im Vergleich zweier Aminosäuren beide eine aliphatische Struktur aufweisen, sich jedoch hinsichtlich ihre Hydrophobizität und Ladung unterscheiden. So lässt sich sagen, dass eine kurze Distanz zwischen zwei Aminosäuren auf stark ähnliche physikalisch-chemische Eigenschaften deutet. Ist die Distanz hingegen größer, so nimmt entsprechend die Ähnlichkeit ab. Dasselbe gilt auch für die einzelnen Anziehungskräfte bzw. Abstoßungskräfte untereinander. Letztendlich werden die räumlichen Zustände, welche eine Peptidkette auf dem Weg der Proteinfaltung durchläuft, wird durch die physikalisch-chemischen Eigenschaften der Aminosäuren und durch deren

Anordnung bzw. deren räumlicher Distanz zueinander bestimmt [5][6]. Der hydrophobe Charakter einer Aminosäure spielt eine weitere wichtige Rolle bei der Betrachtung der intermolekularen Wechselwirkungen zwischen einer Aminosäure und ihrem Lösemittel. So beschreibt die Hydrophobizität das Löslichkeitsverhalten einer Aminosäure (bzw. allgemein: einer Chemikalie) in Wasser und in wasserähnlichen Substanzen (z.B. Ethanol). Ist eine Aminosäure nur sehr gering oder gar nicht in einem solchen Medium löslich dann gilt diese Aminosäure als hydrophob. Bildet eine Aminosäure hingegen Wasserstoffbrückenbindungen mit dem Lösemittel aus, so ist diese gut wasserlöslich und wird als hydrophil bezeichnet. Über diese einzelnen Wasserstoffbrückenbindungen bzw. über die Ausbildung von Wasserstoffbrücken kann die Interaktion einer Aminosäure mit ihrer Umgebung und dem Protein verallgemeinert werden. Hydrophobe Aminosäuren befinden sich häufiger an der Proteinoberfläche, hydrophile Aminosäuren hingegen sitzen eher im Proteininneren. Eine Theorie besagt, dass bei der Proteinfaltung die Faltungsnuclei von hydrophoben Aminosäuren gebildet werden.

Theorie des hydrophoben Effektes. Proteine verhalten sich im wässrigen Medium stabiler, wenn deren hydrophoben Seitenketten im Proteininneren lagern, statt an der Proteinoberfläche in Berührung zu stehen. Denn Wassermoleküle wechselwirken stärker untereinander als mit hydrophoben Seitenketten eines Proteins. Dies hat zur Folge, dass die Polypeptidkette unter Bildung eines kompakten *Molten Globule* in sich zusammen fällt. Dabei nimmt die Entropie der Polypeptidkette im Wechsel vom nicht-nativen zum nativen Zustand ab. Im Gegensatz dazu nimmt die Entropie des Wassers bzw. des wasserähnlichen Lösemittels erheblich zu, denn die Wassermoleküle bilden Käfige um die hydrophoben Gruppen aus, welche während des Faltungsvorganges zerstört werden, des Weiteren werden Wassermoleküle freigesetzt, die vorher mit den Seitenketten verbunden waren. Folglich nimmt die Gesamtenergie des Systems zu. Dieser Anstieg stellt die größte Triebkraft für die Proteinfaltung dar. Die meisten polaren Seitenketten bleiben an der Proteinoberfläche und treten in Kontakt mit den Wassermolekülen, während die unpolaren Seitenketten ins Innere des Proteins gezogen werden. Die Abschnitte des Polypeptidrückgrates, die im Proteininneren lokalisiert sind, können sich mit Hilfe von Wasserstoffbrückenbindungen stabilisieren. Dadurch können sich Sekundärstrukturen ausbilden. Sozusagen trägt die hydrophobe Natur des Proteininneren zur Stabilität von Helices und Faltblättern bei.

Des Weiteren tendieren Aminosäuren, aufgrund ihrer Hydrophobizität, dazu, als Helixbildner, Faltblattbildner oder Turnbildner (Schleifenbildner) zu fungieren. Beispielsweise zählen Aspargin, Aspartat, Prolin und Serin zu den Turnbildnern. Alle vier Aminosäuren gelten als hydrophil. Eine allgemeine Verteilung der einzelnen Aminosäuren in den Sekundärelementen zeigen die nachfolgende Tabelle 2 und Abbildung 18. Sie gibt die konformationelle Präferenz, d.h. das Maß für die Neigung eine Helix, ein Faltblatt oder einen Turn zu bilden, an.

Laut Chou&Fasman gilt: Sind vier von sechs Aminosäuren Helix-bildend, bzw. drei von fünf Faltblatt-bildend, dann ist dieses eine Nukleationsstelle für die jeweilige Sekundärstruktur, von der aus in beiden Richtungen vorgeschritten wird bis eine Tetrapeptid-Sequenz erreicht ist, welche einen Helix- bzw. Faltblattbrecher darstellt [15].

Letztendlich müssen zur Beschreibung des energetischen Zustandes einer Aminosäure in einer Aminosäuresequenz bzw. in einer Peptidkette alle Aminosäure-Aminosäure- und Aminosäure-Lösemittel-Wechselwirkungen erfasst werden.

Tabelle 2 Tabellarische Verteilung der Aminosäuren ein bestimmtes Sekundärstrukturelement zu bilden
[Quelle: nach Vorlage von [15]]

AS	P(α)	P(β)	P(Turn)
Alanin	142	83	66
Arginin	98	93	95
Aspartat	101	54	146
Aspargin	67	89	156
Cystein	70	119	119
Glutamat	151	37	74
Glutamin	111	130	98
Glycin	57	75	156
Histidin	100	87	95
Isoleucin	108	160	47
Leucin	121	130	59
Lysin	114	74	101
Methionin	145	105	60
Phenylalanin	113	138	60
Prolin	57	55	152
Serin	77	75	143
Theronin	83	119	96
Tryptophan	108	137	96
Tyrosin	69	147	114
Valin	106	170	50

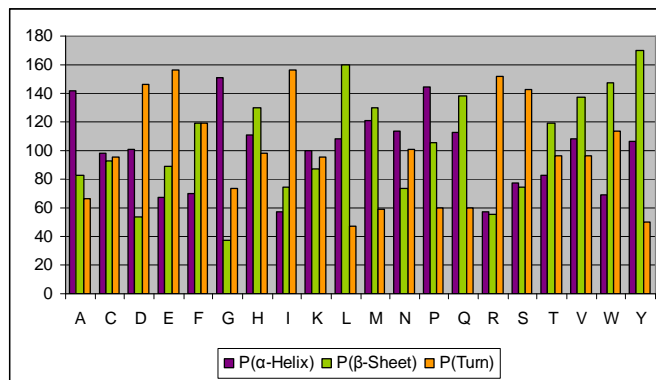


Abbildung 18 Diagramm über die Verteilung bzw. Tendenz der Aminosäuren bestimmte Sekundärstrukturen zu bilden [Quelle: nach Vorlage von [15]].

2.2 Berechnung von Energieprofilen globulärer Proteine

Wie in 2.1 beschrieben, ist der hydrophober Charakter einer Aminosäure von äußerster Relevanz bei der Betrachtung der Proteinfaltung. Aus der Hydrophobizität gegenüber dem Lösemittel und dem Proteins bzw. dem Ausbilden von Wasserstoffbrückenbindungen zum Lösemittel und zum Protein lässt sich ein Innen/Außen-Kriterium $g(i)$ entwickeln. Ist $g(i) = 1$, d.h. die Aminosäure (i) (AS(i)) befindet sich *innen*, wenn für AS(i) gilt:

$$|\vec{c}_\alpha - \vec{c}| \leq 5 \vee (\vec{c}_\alpha - \vec{c}_\beta)(\vec{c}_\alpha - \vec{c}) \leq 0.$$

Formel 2 Innen/Außen-Kriterium einer Aminosäure(i) -
Bedingung so dass für AS(i) gilt: $g(i)=1$ [6].

Dabei entspricht c_α der Position des c_α -Atoms, c dem Schwerpunkt der c_α -Atome aller Aminosäuren innerhalb einer angenommenen Kugel mit dem Radius von 10\AA (Zentrum des Inneren) und c_β der Position des c_β -Atoms.

Ist diese Bedingung nicht erfüllt ist, so befindet sich die Aminosäure (i) *außen*, folglich gilt $g(i) = 0$.

In der nachfolgenden Tabelle sind die Aminosäure-Häufigkeiten von einem 2500 globuläre Proteinen umfassenden Sets dargestellt [6].

Tabelle 3 Die Innen/Außen-Verteilung von Aminosäuren [Quelle: nach Vorlage von [6]].

Aminosäure	innen	außen	Aminosäure	innen	außen
Cys	4582	1016	His	6419	3366
Ile	20370	4141	Gly	16698	14326
Ser	12576	10411	Asp	10001	14327
Gln	7373	7752	Leu	30615	7107
Lys	9285	15193	Arg	11327	10441
Asn	8225	8928	Trp	4001	1193
Pro	9135	9423	Val	23562	6551
Thr	12537	9622	Glu	11165	18091
Phe	13353	2813	Tyr	11228	3529
Ala	22725	11052	Met	7003	1723

Der Zustand einer Aminosäure kann durch eine Variable gekennzeichnet werden. Anhand der Boltzmannverteilung lassen sich nun die Lösungsenergien berechnen.

$$e'_{i0} = -k_B T \ln \left(\frac{n_{in}}{n_{out}} \right)$$

**Formel 3 Boltzmannverteilung
zur Berechnung der
Lösungsenergie einer AS [6].**

Hierbei stellt k_B die Boltzmannkonstante, T die Temperatur und n_{in} bzw. n_{out} die Zustände der Aminosäure (i) dar.

Um alle Aminosäure-Aminosäure-Wechselwirkungen innerhalb eines Proteins beschreiben zu können, wurden Kontaktstatistiken erstellt, anhand derer es möglich ist die Wechselwirkungsenergien abzuleiten. In dieser Statistik wurden alle Kontakte zwischen Aminosäure i und Aminosäure j ausgezählt und als n_{ij} definiert. Die Wechselwirkungen innerhalb eines Proteins lassen sich durch folgende Gleichung beschreiben [5]:

$$e^*_{ij} = -k_B T \ln \left(\frac{n_{ij}}{N_{contact} p_i p_j} \right)$$

**Formel 4 Berechnung der Aminosäure-
Aminosäure-Wechselwirkungsenergien [6].**

$N_{contact}$ beschreibt als Konstante alle beobachteten Kontakte zwischen zwei Aminosäuren aller verwendeten Proteine. Die relativen Wahrscheinlichkeiten p_i und p_j geben die Wahrscheinlichkeit an, dass eine zufällig gewählte Aminosäure eines Proteins der Aminosäure (i) entspricht [6].

Um nur die Energiedifferenzen zu betrachten, muss die obige Gleichung entsprechend umgeformt werden:

$$e_{i0} = \left(\frac{1}{i} \alpha_i \right) e'_{i0}$$

Formel 5 Berechnung der Energiedifferenz [6].

Hierbei beschreibt α_i die Anzahl, mit der die Aminosäure i in Kontakt mit Aminosäure j steht. Dabei wird der Kontakt von AS(i) zu AS(j) durch einen Abstand r_{ij} , welcher kleiner als 8Å beträgt, bedingt. Die Voraussetzung für die Ermittlung eines Abstandes r_{ij} ist das wählen eines räumlichen Bezugspunkts. Dabei werden die Seitenkettenpositionen im Raum interessant, denn diese werden aufgrund der Koordinaten ihrer c_β -Atome, welche stark an das Rückgrat gebunden sind, beschrieben [5][6].

Dementsprechend gilt ein vereinfachtes Paarpotential:

$$e_{ij} = e_{i0} + e_{j0} + e^*_{ij}$$

Formel 6 Berechnung des vereinfachten Paarpotentials von AS(i) und AS(j) [6].

Ausgehend von Formel 6 kann nun für jede Aminosäure innerhalb des Proteins die jeweilige Energie und ein Energieprofil für das gesamte Protein kalkuliert werden. Ein Energieprofil stellt im Prinzip die Auftragung der einzelnen Aminosäureenergien entlang der Aminosäuresequenz dar und kann im zweidimensionalen Raum abgebildet werden [5].

Die Gesamtenergie eines gefalteten globulären Proteins setzt sich durch die Wechselwirkungen bedingten Energien der einzelnen Aminosäuren zusammen. So lässt sich die Gesamtenergie folgendermaßen berechnen:

$$E_{ges} = \sum_{ij} e^*_{ij} f(r_{ij}) + \sum_i e'_{i0} g(i)$$

Formel 7 Berechnung der Gesamtenergie eines globulären Proteins [6].

Dabei entspricht e^*_{ij} der Wechselwirkung zwischen Aminosäure (i) mit Aminosäure (j). Die Abstandsfunktion zwischen Aminosäure (i) und Aminosäure (j) wird von $f(r_{ij})$ beschrieben, sie relativiert die Wechselwirkungen. Die Interaktion einer Aminosäure (i) mit dem umgebenen Lösemittel wird von e'_{i0} erfasst und der Zustand einer Aminosäure (i) wird anhand von $g(i)$ veranschaulicht [5][6].

Die Tools auf dem ePros Server, auf den Seiten der HS Mittweida, sind unter anderen in der Lage EPs von Sequenzen bzw. bekannten Strukturen zu erstellen und zwei EPs miteinander zu alignen. In der nachfolgenden Abbildung 19 wurde aus einer Struktur -2Q37- der OHCU Decarboxylase ein EP kalkuliert. Zusätzlich zum EP wird auch die vorhandene dreidimensionale Struktur dargestellt. Wird nun im EP eine AS oder ein gesamter Bereich markiert, so wird diese AS bzw. dieser Bereich in der dreidimensionalen Abbildung der Struktur entsprechend eingefärbt.

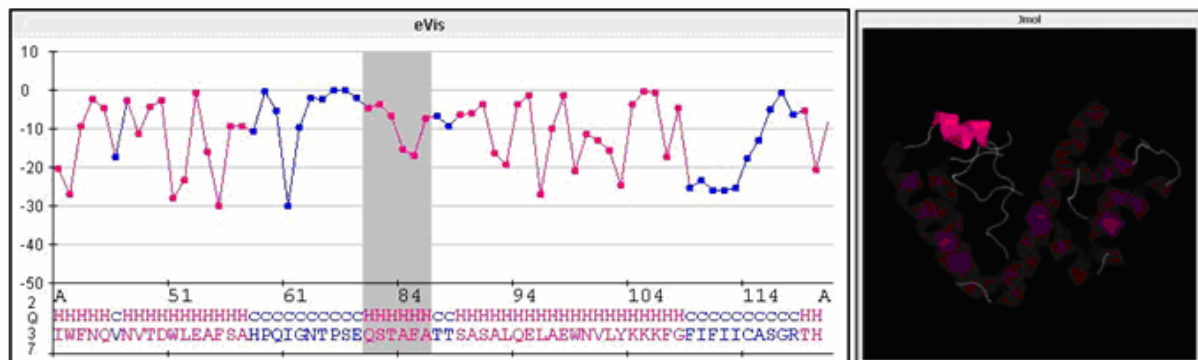


Abbildung 19 Links: Ausschnitt des Energieprofil (EP) der Struktur 2Q37 (PDB-ID) im Visualisierungstool eVis auf dem ePros-Server der HS Mittweida. Die EP-Darstellung zeigt ein Diagramm, wobei die Energie (10 = hohe Energie, -50 = niedrige Energie) zum einen gegen die AS-Position aufgetragen wird und zum anderen gegen die AS entsprechende Sekundärstruktur. Die Farbgebung richtet sich nach den Sekundärstrukturelementen, hierbei werden helicale Strukturen rosa und coil Strukturen blau eingefärbt. Der grau unterlegte Bereich entspricht einer Helix (siehe rechtes Bild). Rechts: 3D-Struktur von 2Q37 dargestellt in Jmol. Die rosa hervorgehobene Helix entspricht dem im EP grau unterlegten Bereich [25].

3. Analyse von 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase

Nierensteine und Gicht resultieren aus einer zu hohen Konzentration an Harnsäure im menschlichen Blut, einer Hyperurikämie. Die Harnsäure stellt das Endprodukt des menschlichen Purinkatabolismus dar. Sie ist eine Schlüsselkomponente in dem Metabolismus von Purinen und wird in vielen Organismen zu (S)-Allantoin abgebaut. In dem kompletten Harnsäure Abbau sind drei enzymatische Reaktionen involviert. Zunächst wird Harnsäure oxidiert. diese Reaktion wird katalysiert von der Urate Oxidase und liefert das Zwischenprodukt 5-Hydroxyisourate (HIU), welches anschließend mittels der HIU Hydrolase zu 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline (OHCU) reagiert. OHCU wird stereochemisch decarboxyliert und führt schließlich zu (S)-Allantoin und Kohlenstoffdioxid [16].

Im menschlichen Organismus fehlt das Enzym Urate Oxidase, sodass Harnsäure das Endprodukt des Purinkatabolismus im Menschen darstellt [22]. Harnsäure wird in der Leber gebildet, ins Blut abgegeben und über die Niere ausgeschieden. Es ist bis zu zehnfach höher im Blut konzentriert ist als in anderen Primaten. Gicht entsteht, wenn die Harnsäurekonzentration im Blut auf das 15-25fache erhöht. Aufgrund der schlechten Löslichkeit der Harnsäure und ihrer Salze können Nierensteine und schmerzhaft Entzündungen resultieren [3]. Doch warum verlor ein so erfolgreich entwickeltes Wesen wie der Mensch die Fähigkeit Harnsäure abzubauen? Eine Vermutung besagt, dass Harnsäure auch nützlich für den Organismus ist, denn sie gilt als Antioxidant. Daher ist sie in der Lage DNA bedrohende Substanzen zu neutralisieren. Nicht nur beim Menschen fehlt Urate Oxidase, auch andere Primaten, Vögel, Reptilien, Dalmatiner und manche Mikroorganismen haben diese Funktion im Laufe der Evolution verloren [22].

Die meisten Säugetiere, mit Ausnahme des Menschen, anderen Primaten, Vögeln und Reptilien, bauen Harnsäure weiter zu Allantoin ab. Manche andere Tiere führen den Harnsäureabbau weiter bis hin zu Allantoinsäure oder zu Harnstoff und Glyxolat. Doch aufgrund der Tatsache, dass der Purinabbau beim Menschen mit der Abbaustufe Harnsäure endet, welches im Gegensatz zu Allantoin eine sehr schlechte Wasserlöslichkeit aufweist, ist der Mensch anfällig für aus Hyperuricämien resultierende Krankheiten. Durch zu hohes Harnsäure Angebot oder bedingt durch eine gestörte Harnsäureverwertung entstehen kristalline Harnsäureablagerungen im

menschlichen Körper, speziell in den Gelenken. Diese Ablagerungen rufen folglich schmerzhafte Gichtanfälle hervor. Zu stark purinhaltige Nahrung wie beispielsweise Fleisch wirkt sich ebenfalls eher ungünstig auf den Harnsäurehaushalt aus [18].

Doch warum gibt es Organismen, welche Harnsäure zu weiteren, besser löslichen Produkten abbauen können um diese für den Körper toxische Harnsäure unschädlich zu machen? Tatsächlich besitzt der Mensch die Gene, welche die Enzyme zum weiteren Abbau der Harnsäure zu dem gut wasserlöslichen Allantoin codieren. Diese notwendigen Enzyme sind die Urate Oxidase, die HIU Hydrolase und die OHCU Decarboxylase. Doch herauszufinden wie die drei Gene inaktiviert wurden, wird noch eine sehr lange Zeit in Anspruch nehmen.

3.1 Strukturaufklärung und Modellierung des aktiven Zentrums

Anhand der Kristallstruktur von OHCU Decarboxylase des Zebrafisches und basierend auf den Strukturen des Enzyms im Komplex mit zwei Liganden bekamen Ramazzina et al. Einblicke in die Mechanismen und Stereoselektivität der enzymatischen Reaktion von OHCU Decarboxylase des Zebrafisches. Im Paper *The Structure of 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase Provides Insights into the Mechanism of Uric Acid Degradation* wurden ihre Ergebnisse dargestellt [16].

Komplette Struktur von OHCU Decarboxylase. OHCU Decarboxylase ist ein Homodimer. Es handelt sich hierbei um einen Komplex aus zwei identischen Proteinmolekülen. Zwei Polypeptidketten sind in einer 2-Schritt-Achse zueinander ausgerichtet. Jedes Monomer besteht aus 174 Aminosäuren, bildet sich aus 10 α -Helices und ist somit aus zwei Domänen zusammengesetzt. Bei Domäne I handelt es sich um die N-terminale Domäne (Abbildung 20). Sie erstreckt sich von Rest Asp-2 bis Rest His-67 und beinhaltet die ersten fünf α -Helices (α_A bis α_E) (siehe Abbildung 14). Die α -Helices A bis E gelten als eher kurz, da sie zwischen fünf und zehn Reste lang sind [16].

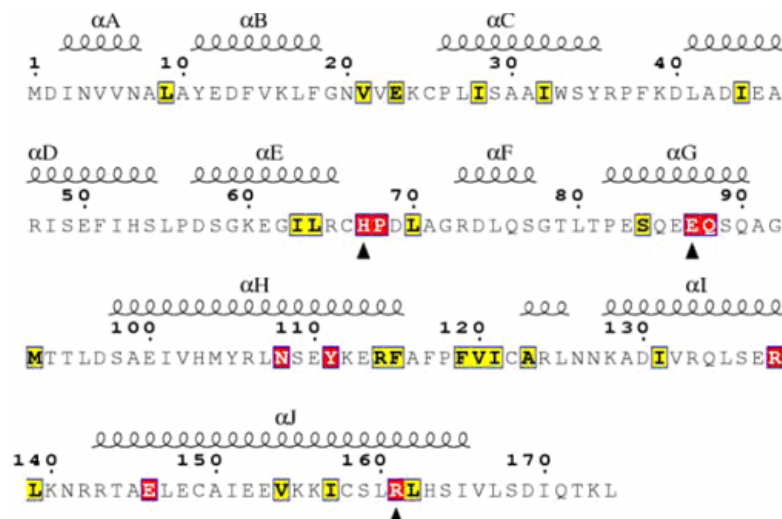


Abbildung 20 Aminosäuresequenz der Zebrafisch OHCU Decarboxylase. Hoch konservierte oder chemisch übereinstimmende Aminosäurereste sind in den roten und gelben Boxen dargestellt und resultieren aus einem multiplen Alignment von 60 Proteinsequenzen verschiedener Organismen. Über der Aminosäuresequenz sind die zehn α -Helices abgebildet und mit A bis J markiert. Die schwarzen Dreiecke zeigen auf die drei vermutlichen an der Katalyse beteiligten Aminosäuren, basierend auf Mutationsanalysen von Ramazzina et al. [16].

Wobei sich die α_C -Helix ziemlich irregulär an ihrem Ende verhält, denn sie schwankt zwischen sieben und zehn Resten. Eine Ausnahme bildet jedoch die 14 Reste lange α_D -Helix [16].

Einen Überblick der Anordnungen von Sekundärstrukturelementen und Strukturmotiven wird in der nachfolgenden Tabelle 4 beschrieben. Die Sequenz des Modelorganismus Zebrafisch, welche in der Protein Daten Bank (PDB) unter dem PDB Code 2o70 geführt wird, beinhaltet neben den zehn α -Helices, zwei 3.10-Helices, sechs β -turns, wovon drei Wasserstoff gebundene turns sind, und einen γ -turn. Bei einem β -turn handelt es sich um ein Motiv, welches definiert ist für vier aufeinander folgende Reste, die mit i , $i+1$, $i+2$, $i+3$ bezeichnet werden. Die Distanz zwischen dem C_α -Atom des Restes i und des C_α -Atoms des Restes $i+3$ muss weniger als 7Å betragen und die zwei zentralen Reste dürfen keine Helices sein. Ein γ -turn – Motiv ist definiert für drei Reste (i , $i+1$, $i+2$), wenn eine Wasserstoffbrücke zwischen den Resten i und $i+2$ besteht [19].

Tabelle 4 Tabellarische Zusammenfassung der Strukturelemente über die gesamte Sequenzlänge der Zebrafisch OHC Decarboxylase (Kette A). Legende: orange = Wasserstoff gebundener β -turn; blau = nicht Wasserstoff gebundener β -turn; grün = γ -turn; violett = 3.10 Helix; grau = keine strukturellen Daten verfügbar [Datenquelle: [19][20]]

Aminosäuren	Position (Länge)	Sekundär Struktur Element	Motiv
MD	1-2 (2)		
INVVN	3-7 (5)	α -Helix _A	
ALA	8-10 (3)		
YEDFVKLF	11-18 (8)	α -Helix _B	β -turn
GNVVEKC	19-25 (7)		FGNV/ NVVE/ VEKC
PLISAAIWSY	26-35 (10)	α -Helix _C	
RPFKD	36-40 (5)		
LADIEARISEFIHS	41-54 (14)	α -Helix _D	
LP	55-56 (2)		
DSGKEGILR	57-65 (9)	α -Helix _E	
CHPDLAG	66-72 (7)		β -turn DLAG
RDLQS	73-77 (5)	α -Helix _F	
GTLT	78-81 (4)		
PESQEEQSQ	82-90 (9)	α -Helix _G	
AGMTTLD	91-97 (7)		β -turn GMTT MTTL
SAEIVHMYRLNSEYKERF	98-115 (18)	α -Helix _H	
GFPFVIC	116-122 (7)		γ -turn
ARL	123-125 (3)	3.10-Helix	ICA
NN	126-127 (2)		
KADIVRQLSER	128-138 (11)	α -Helix _I	
LKNR	139-142 (4)		
RTAELECAIEEVKKICSLRLHSI	143-165 (23)	α -Helix _J	
VLSDIQTKL	166-174 (9)		

Die C-terminale Domäne (Domäne II) beinhaltet die Reste von Pro-68 bis Ile-165 und folglich die restlichen fünf α -Helices (α_F bis α_J) (siehe Abbildung 20). Am kürzesten ist die α_F -Helix mit vier Resten. Die α_H -Helix sowie die α_J -Helix sind mit 18 beziehungsweise 22 Resten die längsten Helices im gesamten Enzym. Alle α -Helices sind durch kurze Loops miteinander verbunden [16].

Das Segment 66-72, welches in einer Art erweiterter Konformation liegt, verbindet Domäne I und Domäne II miteinander. Ein ebenfalls hauptsächlich in der erweiterten Konformation liegender Strand ist präsent in Domäne II, indem er die α_H -Helix und α_J -Helix verknüpft [16].

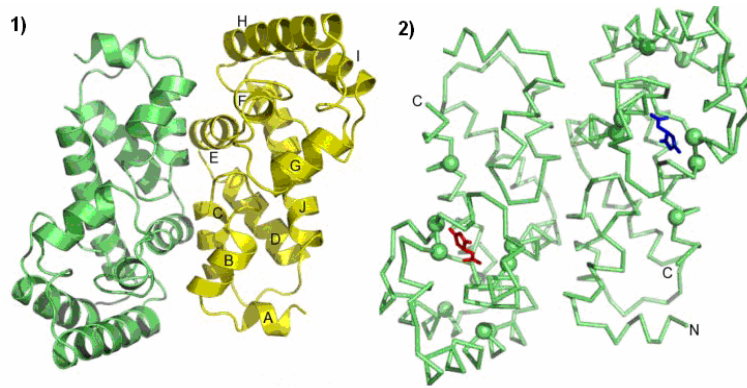


Abbildung 21 1) 3D-Darstellung des Dimers OHCU Decarboxylase vom Zebrafisch. Erkennbar sind die 2-Schritt-Achsen (siehe auch Abbildungen 16). Die α -Helices sind von A bis J beschriftet. 2) 3D-Darstellung von C_{α} Kette des Enzym Dimers mit gebundenem Allantoin (rot bzw. blau) in den beiden aktiven Zentren. N und C bezeichnet den N- und C-Terminus [16].

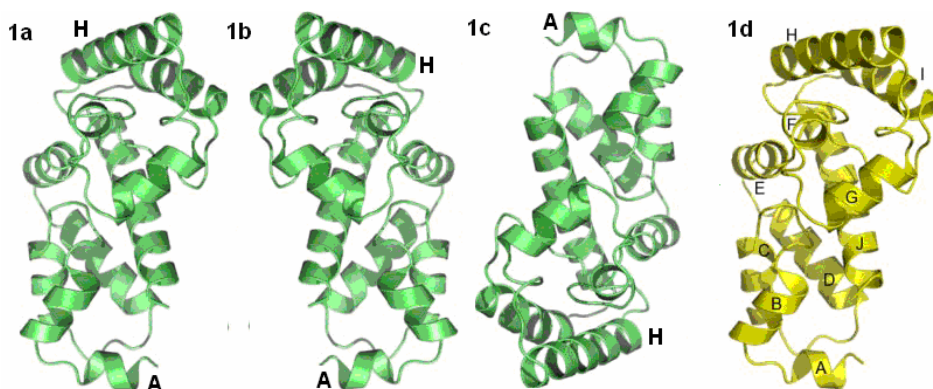


Abbildung 22 Bildung der 2-Schritt-Achse des ist abgebildet von 1a bis 1c. 1d zeigt das zweite Monomer (gelb) [16].

In der Abbildung 22 ist die Bildung der 2-Schritt-Achse dargestellt. 1a ist der Ausgangszustand des ersten Monomers. Dieser Monomer ist identisch mit dem zweiten Monomer (gelb, 1d). Das erste Monomer wird im ersten Schritt vertikal gespiegelt (1b) und im zweiten Schritt noch einmal horizontal gespiegelt (1c). Sodass der fertige Dimer aus einer 2-Schritt-Achse besteht (siehe Abbildung 21).

Die Aminosäuresequenzen der OHCU Decarboxylase weist keinerlei Homologie zu anderen bekannten Aminosäuresequenzen auf. Dies wurde in strukturellen Suchen von Ramazzina et al. belegt, da keine Ähnlichkeiten zu anderen Proteinen gefunden

wurden. Demnach besitzt das Enzym OHCU Decarboxylase möglicherweise eine zuvor nie gesichtete helicale strukturelle Faltung [16][17].

In einem Monomer befinden sich 21 potentiell positiv geladene und 25 potentiell negativ geladene Aminosäurenreste. Die meisten dieser Reste wurden auf der Oberfläche des Moleküls aufgespürt, mit Ausnahme von Glu-146, Arg-138, Glu-23, Arg-161 und Glu-87. Die geladenen Seitenketten dieser Reste zeigen in das Innere des Proteinmoleküls. Aufgrund ihrer Ladung sind sie in elektrostatische Interaktionen verwickelt. Jedes Monomer enthält auch fünf Cystein-Reste. Cystein besitzt Schwefelatome und kann somit Disulfidbrücken ausbilden, welche hier aber nicht geschieht. Im dimerischen Stadium des Proteins bestehen Wechselwirkungen zwischen α_C -Helix des einen Monomers und α_E -Helix des anderen Monomers. Ebenfalls auftretende Interaktionen gibt es zwischen der Loop verbindenden α_B -Helix zu α_C -Helix beider Monomere. Die Fläche der Dimerisation beträgt ca. 2,3 Å² pro Monomer. Insgesamt beinhaltet sie 14 Wasserstoffbrückenbindungen [16].

Struktur des Allantoin -OHCU Decarboxylase - Komplexes und des aktiven Zentrums. Da es sich bei OHCU Decarboxylase um einen Dimer handelt, existieren zwei aktive Zentren. Beide werden durch jeweils einen Liganden belegt. Dabei entstehen keine Wechselwirkungen, da beide Spalten räumlich weit voneinander entfernt liegen (Abbildung 21: 2)). Im ungebundenen Zustand des aktiven Zentrums, das heißt wenn kein Ligand zur Verfügung steht, füllen jeweils fünf Wassermoleküle den Raum der zwei Spalten aus [16]. Im aktiven Zentrum bestehen keine Notwendigkeiten und Interaktionen mit CoFaktoren, wie beispielsweise Metallen [17]. Das Allantoin ist in der katalytischen Spalte des Enzyms gebunden. Diese befindet sich nahe der äußeren Oberfläche des Proteins und zwischen α_G -Helix und dem Streches, welcher α_H -Helix - α_I -Helix und α_E -Helix - α_F -Helix verbindet. Das katalytische Zentrum ist am Rand der Domäne II eines jeden Monomers lokalisiert und steht im engen Kontakt zu Domäne I, welche partiell die Oberfläche des katalytischen Zentrums berührt. Hydrophile und Hydrophobe Reste berühren die interne Oberfläche der katalytischen Spalte, deren Spitze mittels der Seitenketten von Glu-87 und Arg-161 abgeschlossen wird. Hierbei wird Glu-87 durch Arg-161 neutralisiert. Arg-161 interagiert zugleich mit Glu-23, welches von His-67 neutralisiert wird. Demnach ist eine Anordnung von vier alternierenden Ladungen (Glu-87-Arg-

161-Glu-23-His-67) im enzymatischen aktiven Zentrum präsent. Den hydrophoben Teil der Oberfläche der katalytischen Spalte formen die Seitenketten von Leu-70, Val-120, Ile-157 und Ala-123. Kim et al. fanden heraus, dass (S)-Allantoin in eine hydrophobe Umgebung eingebettet ist, welche jedoch mit hydrophilen Resten auf der Oberfläche der Spalte bedeckt ist [16][17]. Spezifische Interaktionen werden durch die konservierten Reste (His-58, Glu-80, Arg-153) herbeigeführt, weil deren Seitenketten direkt mit dem gebundenen Allantoin agieren. Glu-80 und Arg-153 leisten darüber hinaus auch strukturelle Funktionen im aktiven Zentrum. Die COOH-Gruppe der Seitenkette von Glu-80 bildet zweisträngige Wasserstoffbrücken mit einer Länge von 3 Å zu der Seitenkette von Arg-153 aus. Dabei entsteht eine Überbrückung des N-Terminus (Domäne I) und des C-Terminus (Domäne II). Die daraus resultierende Konformationserweiterung dieser Seitenketten dient als eine Art Klappe zur Abgrenzung des Allantoin zum Lösemittel [17].

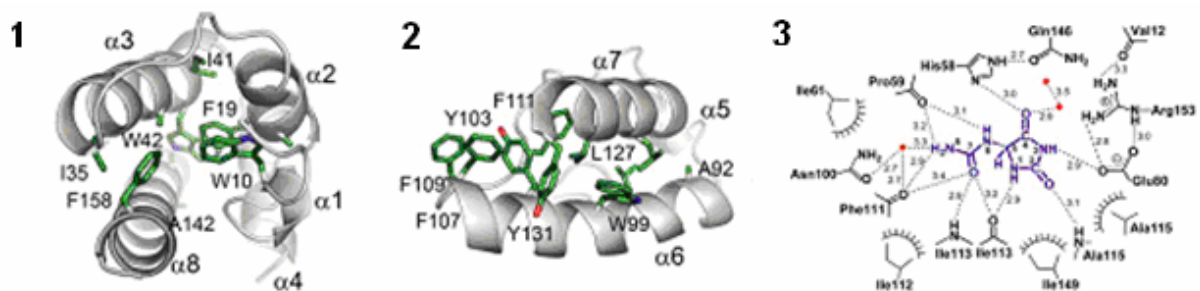


Abbildung 23 1) Darstellung der N-terminalen Domäne eines Monomers. Der Kern besteht aus konservierten hydrophoben Resten. [16] 2) Interaktionen, meist herbeigeführt durch hydrophobe Reste, werden im C-Terminus dargestellt [16]. 3) Wechselwirkungen zwischen gebundenem (S)-Allantoin und Resten des aktiven Zentrums. Die gestrichelten Linien zeigen mutmaßliche Wasserstoffbrücken mit zugehörigen interatomaren Abständen. Van der Waals Kräfte ($<4,0$ Å) werden durch Bögen mit Strichen dargestellt [17].

3.2 Funktion und Einordnung in den Purinkatabolismus

Die OCHU Decarboxylase katalysiert den letzten Schritt im Purinkatabolismus, die Umwandlung von 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline zu dem optisch aktiven S-Allantoin [16][21], wobei S-Allantoin rechtsdrehend ist (Abbildung 24, rote Pfeile).

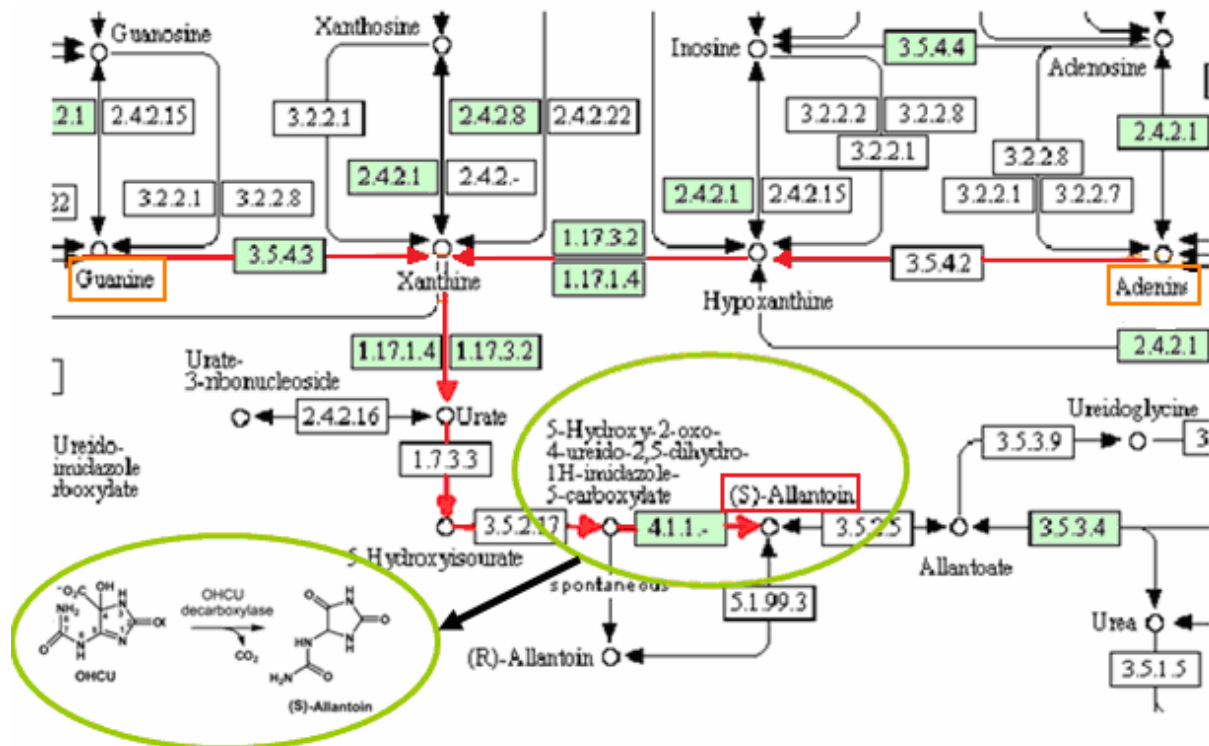


Abbildung 24 Ausschnitt des Purinkatabolismus. Rote Pfeile markieren den Abbauweg der Purine Guanin und Adenin (orange eingerahmt): Umwandlung von Xanthin zu (S)-Allantoin (rot eingerahmt). Im grünen Kreis findet die Umwandlung von OCHU zu (S)-Allantoin statt, welche von der OCHU Decarboxylase katalysiert wird. Des Weiteren sind die EC-Nummern der beteiligten Enzyme (in Rechtecken gerahmt) angegeben. [Bildquelle: Seiten der KEGG Datenbank]

Die Purine Adenin und Guanin sind Doppelringbasen und Bausteine der Nukleinsäuren, welche die DNA bilden. Sie werden über Desaminierung und der Oxidation des Puringerüsts zu Hypoxanthin beziehungsweise Xanthin, welches anschließend zu Harnsäure destruiert wird, abgebaut. In den Oxidationsreaktionen wird jeweils eine, aus molekularem Sauerstoff stammende Oxo-Gruppe eingeführt [10]. Die Harnsäure wird anschließend mittels des Enzyms Uratoxidase unter Verwendung von Sauerstoff und Wasser zu 5-Hydroxyisourate (HIU) und Wasserstoffperoxid (H_2O_2) oxidiert. Das toxische H_2O_2 wird anschließend von

Peroxidasen entfernt. Mittels der HIU Hydrolase wird HIU zu OHCU hydrolysiert. OHCU wird schließlich, unter Mitwirkung von OHCU Decarboxylase, zu (S)-Allantoin stereochemisch decarboxyliert (siehe Abbildung 24).

Mutmaßlicher katalytischer Mechanismus der OHCU Decarboxylase des Modelorganismus Zebrafisch. Wie in der nachfolgenden Abbildung 25 dargestellt, nimmt Histidin58 (His58) an der stereochemische Decarboxylierung von OHCU zu (S)-Allantoin teil. Es findet eine Decarboxylierung des C4-Atoms und eine Protonierung am C5-Atom statt. Das Kohlenstoffatom wird durch Verwendung der Doppelbindung zwischen C5 und N1 und dem daraus resultierenden Elektronenabfluss stabilisiert [17]. So entsteht am C4-Atom eine Doppelbindung zum Sauerstoff. Die Doppelbindung zwischen C4- und C5-Atom wird durch in eine Einfachbindung umgewandelt, da C nur vier Außenelektronen besitzt.

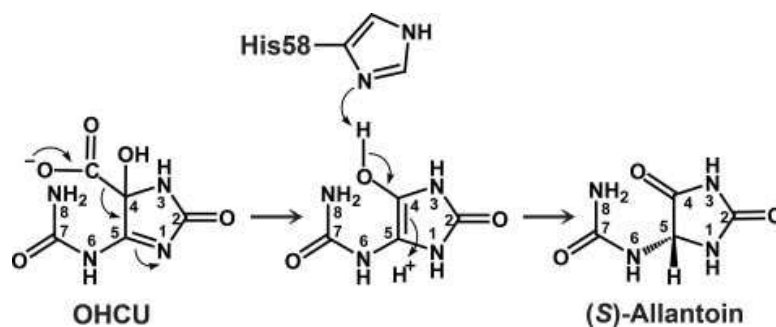


Abbildung 25 Vermutlicher Mechanismus der OHCU Decarboxylase von *A.thaliana* unter der Annahme, dass Histidin (His) im aktiven Zentrum der OHCU Decarboxylase sitzt [17].

Mutmaßlicher katalytischer Mechanismus der OHCU Decarboxylase des Modelorganismus Zebrafisch. Die Carboxylgruppe des Glutamin87 (E87) (Abbildung 26) ist nahe der des Substrates (Abbildung 26, zentral, rötlich, C-Atome von 1-6 nummeriert), wobei der Abstand dieser beiden negativ geladenen Gruppen ca. 3Å beträgt. Der geringe Abstand verursacht womöglich elektrostatische Abstoßungen zwischen den Gruppen. Zwischen dem Imidazol-Ring von His-67 und dem C5-Atom des fünfgliedrigen Ringes des Substrates herrscht ein Abstand von 5Å vor. Dieser kann aber durch eine Rotation um die C α -C β Bindung von His-67 auf bis zu mindestens 3Å verkürzen Die Reste Glutamin87 (E87) und Histidin67 (H67) sind nahe dem Substrat, somit lässt ihre Position vermuten, dass sie eine kritische Rolle

in der Katalyse spielen. Die enzymatische Reaktion könnte auch indirekt durch Arginin161 (R161) und Glutamin 23 (nicht in Abbildung 26 abgebildet) betroffen sein. Denn die zwei geladenen Reste dieser Aminosäuren tragen eventuelle zum Ladungsgleichgewicht der katalytischen Spalte bei [16].

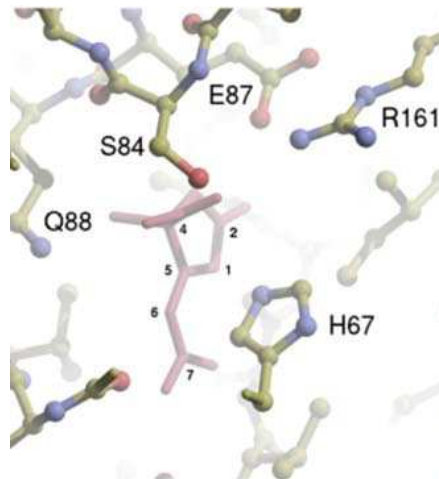


Abbildung 26 Blick auf den möglichen OHCU-OHCU Decarboxylase Komplex im aktiven Zentrum der Zebrafisch-OHCU Decarboxylase. Die Aminosäuren sind im Ein-Letter-Code benannt mit ihrer zugehörigen Position in der AS-Sequenz des Proteins[16].

Die Nähe der Carboxylgruppe von Glutamin87 (E87) könnte eine Destabilisierung des Substrat-Grundzustandes bewirken. Dies würde zu einer erleichterten Freilassung der Carboxylgruppe als CO_2 führen. Anschließend wird ein Enol-Zwischenprodukt gebildet. Nach einer finalen Umordnung entsteht schließlich das Keto-Produkt. Diese Enol-Keto-Tautomerisation kann durch die Anwesenheit von Histidin67 (H67) beeinflusst werden. Denn Histidin67 wird protoniert und befindet sich nahe genug am C5-Atom des Substrates und kann somit ein Proton zum Enol-Zwischenprodukt donieren um letztendlich das finale Keto-Produkt zu bilden. Aufgrund der *re*-Position von Histidin67 zum gebundenen Substrat, wird Allantoin in seiner korrekten (S)-Konfiguration gebildet [16].

In dieser Hypothese von Ramazzina et al spielt Histidin67 die Hauptrolle in der Katalyse. Histidin67 dient während der Reaktion als Protonendonator und kontrolliert die Stereochemie des Produktes. Mit Hilfe einer Mutation am Histidin68 der *B.subtilis*

OHCU Decarboxylase (bei Ramazzina at all, Zebrafisch OHCU Decarboxylase Histidin67) gelang es Kim at all eine Inaktivierung der Katalytischen Fähigkeit des Enzyms festzustellen. Bedingt durch die Inaktivierung wurde eine starke Verlangsamung des Abbaus von OHCU notiert. Mutationen am Glutamin84 (bei Ramazzina at all Glutamin87) und Arginin158 (bei Ramazzina at all Arginin161 (R161)) inaktivierten ebenfalls das Enzym, möglicherweise durch beschädigen der katalytischen Rolle der Enzymaktivität oder durch mindern der Bindungsaffinität des Substrates zum Enzym [17].

Lange Zeit wurde vermutet, dass Harnsäure direkt durch die Urate Oxidase zu Allantoin umgewandelt wurde, wie aus Abbildung 27 zu entnehmen. Doch dies wurde unter anderem durch Ileana Ramazzina und anderen Forschern in mehreren Veröffentlichungen, beispielsweise *The Structure of 2-oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase Provides Insights into the Mechanism of Uric Acid Deradation*, widerlegt.

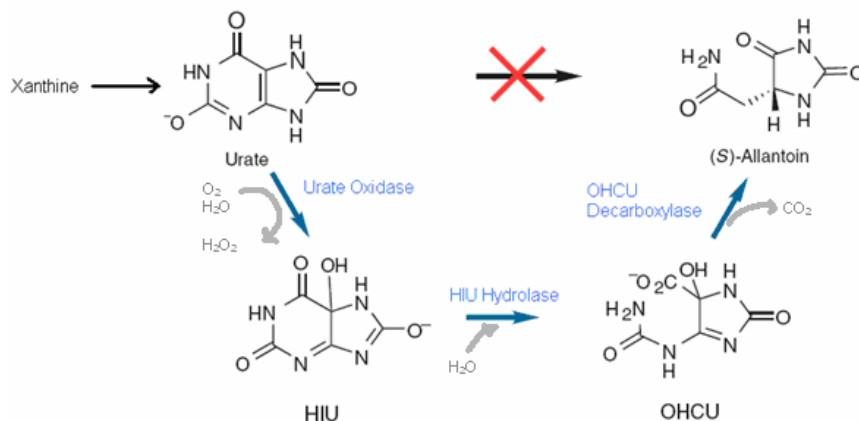


Abbildung 27 Harnsäure Metabolismus. Harnsäure (Urate) wird nicht wie lange angenommen durch einen einzelnen Schritt zu Allantoin abgebaut sondern mithilfe der Enzyme HIU Hydrolase und OCHU Decarboxylase [22].

Allantoin (siehe Anhang, Tabelle 9) besitzt ein chirales Kohlenstoffatom, daher eine optische Isomerie und kann somit in (R)- und (S)-Form auftreten. Das (S)-Allantoin kann sich in manchen Organismen weiterführenden enzymatischen Reaktionen unterziehen. Der vollständige Abbau von (S)-Allantoin liefert Kohlenstoffdioxid (CO₂) und Ammoniak (NH₃). Reagiert jedoch das OHCU spontan zu einer racemischen Mischung aus (R)-Allantoin und (S)-Allantoin so entspricht dieser Vorgang einer Reaktion ohne enzymatische Beteiligung. Da (R)-Allantoin eine gespiegelte

Stereochemie gegenüber (S)-Allantoin aufweist, kann es nicht weiter zu Allantoate und folglich auch nicht zu weiterführenden Produkten abgebaut werden. Denn das an der Umwandlung von (S)-Allantoin zu Allantoate beteiligte Enzym Allantoinase [6] ist stereospezifisch und somit an die (S)-Form des Allantoins gebunden [3][22].

4. Betrachtung der Familie OHCU Decarboxylase auf der Ebene der Sequenzen

Mit Hilfe von ClustalW wurden für vier unterschiedliche Organismengruppen MSAs erstellt. Die Enzym Familie 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase wurde in zwei Hauptgruppen unterteilt: Eukaryonten und Prokaryonten. Diese beiden Hauptgruppen, welche je 40 Sequenzen umfassen, wurden erneut in zwei Gruppen unterteilt. Die Gruppe der Eukaryonten gliedert sich in *Metazoa* und *Fungi* (Pilze), die Prokaryonten werden in *Actinobacteria* und *Proteobacteria* differenziert. Jeder Untergruppe besitzt je 20 Sequenzen. Eine Gruppierung in 2 bzw. 4 repräsentative Gruppen aus über 400 Sequenzen und mehr als 200 Spezies hat den Vorteil, dass man einen groben Überblick über die Sequenzen (bzw. in Punkt 5 Energien) erhält und erste Aussagen über Unterschiede und Übereinstimmungen zu treffen, sowie deren evolutionäre Entwicklung aufzuklären.

Die Gruppe der Eukaryonten umfasst unter anderem Pilze, Tiere, Pflanzen, und Ciliaten. Zu den Prokaryonten zählen vor allem Bakterien. Eukaryotische Zellen unterscheiden sich nicht nur in ihrer Größe, die eukaryontische Zelle ist etwa 10mal größer als die wenige μm große prokaryontische Zelle, sondern auch in ihrer stofflichen Zusammensetzung und in ihren Feinstrukturen von den prokaryontischen Zellen. Der eindeutigste Unterschied besteht darin, dass die eukaryontische Zelle einen von einer Membran umgebenen Zellkern (Nukleus) und die prokaryontische Zelle keinen Zellkern, sondern eine kernähnliche Struktur (Nukleoid) besitzt [1].

In 4.1 werden nun vier MSAs vorgestellt, welche unter anderem erste Rückschlüsse auf Konservierungen zulassen. 100%-ige Konservierungen erkennt man im Alignment daran, dass die jeweilige Aminosäure an einer bestimmten Position in jeder AS-Sequenz auftritt.

4.1 Multiple Alignments basierend auf Sequenzen

Das Farbschema nach Clustal, welches in der nachfolgenden Abbildung 28 dargestellt ist, gilt für alle folgenden multiplen Alignments auf Sequenzebene.

Clustal X Default Colouring		
Residue at position	Applied Colour	{ Threshold, Residue group }
A,I,L,M,F,W,V	BLUE	{+60%, WL,VIMAFCHP}
R,K	RED	{+60%,KR},{+80%, K,R,Q}
N	GREEN	{+50%, N},{+85%, N,Y}
C	BLUE	{+60%, WL,VIMAFCHP}
C	PINK	{100%, C}
Q	GREEN	{+60%,KR},{+50%,QE},{+85%,Q,E,K,R}
E	MAGENTA	{+60%,KR},{+50%,QE},{+85%,E,Q,D}
D	MAGENTA	{+60%,KR},{+85%, K,R,Q},{+50%,ED}
G	ORANGE	{+0%, G}
H,Y	CYAN	{+60%, WL,VIMAFCHP},{+85%, W,Y,A,C,P,Q,F,H,I,L,M,V}
P	YELLOW	{+0%, P}
S,T	GREEN	{+60%, WL,VIMAFCHP},{+50%, TS},{+85%,S,T}

Abbildung 28 ClustalW Farbschema in JalView [Quelle: [23]]

Eukaryonten. Metazoa.

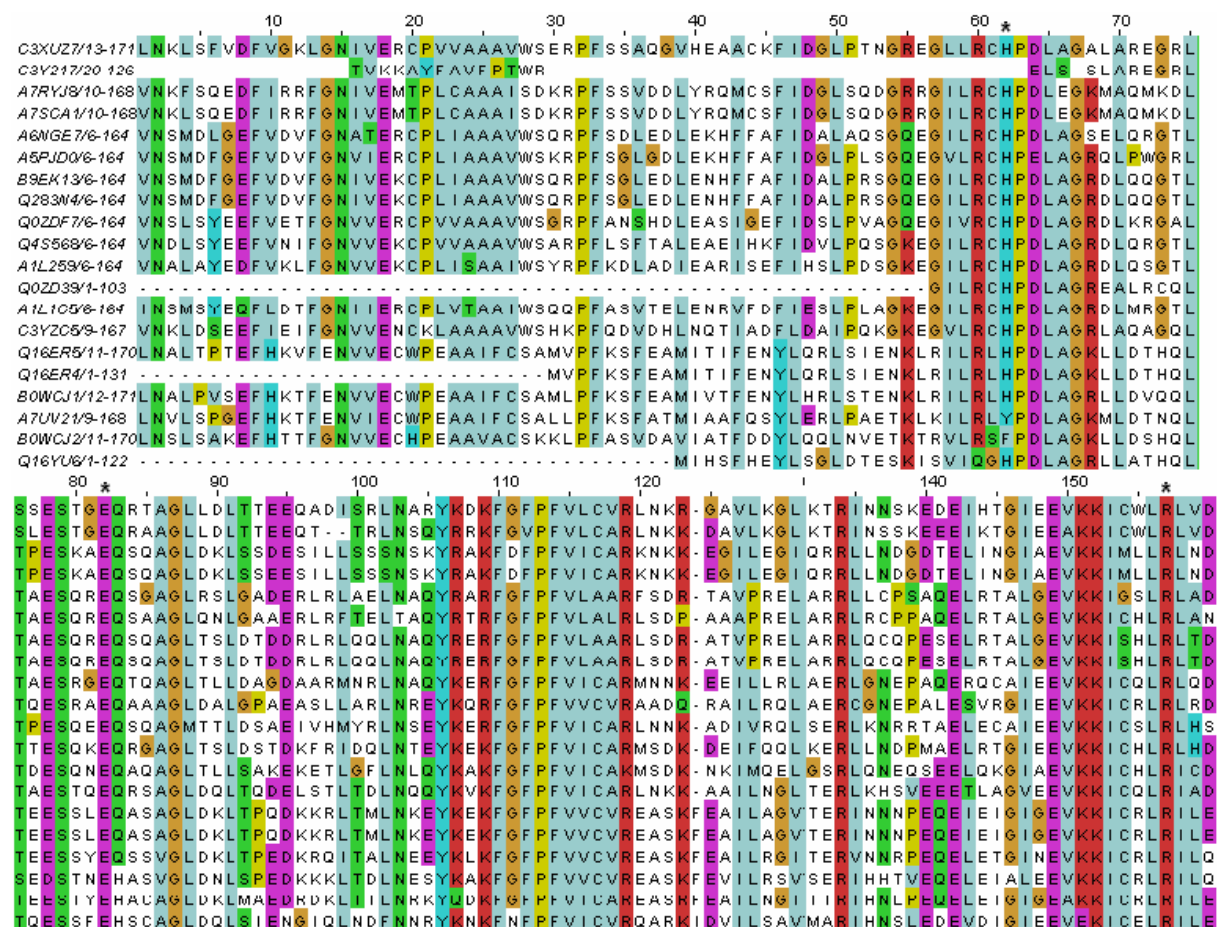


Abbildung 29 JalView Ansicht des *Metazoa-OHCU-Dec-MSA*, welches mittels ClustalW erstellt wurde. Mit schwarzem Stern markierte AS sind laut [16] vermutlich an Katalyse beteiligt [23].

Das MSA der Gruppe *Metazoa* (siehe Abbildung 29) enthält 20 Sequenzen mit einer durchschnittlichen AS-Länge von 149. Aus dem MSA und der *Conservation-Sequence* (nicht abgebildet) ist erkennbar, dass 18 hochkonservierte (zu 100% konservierte) Aminosäuren existieren (siehe Tabelle 5). Dazu zählen beispielsweise die AS Leucin (L) an Position 75 sowie 91, und die AS Tyrosin (Y) an Position 106. Des Weiteren gibt es eine Reihe isofunktioneller Austausche, d.h., dass eine AS an dieser Position der Sequenz durch eine andere, aber in ihrer Struktur und physikochemischen Eigenschaft ähnlichen AS, ausgetauscht wurde und somit wird die Funktion und die Struktur des Enzyms nicht beeinträchtigt. Ein isofunktioneller Austausch befindet sich an Position 64 im Alignment. An dieser Position wird die Asparginsäure (D) zweimal mit Glutaminsäure (E) ausgetauscht. Beide AS sind sauer, folglich gibt es keine Einschränkung für die Funktion des Enzyms. Ebenfalls ein isofunktioneller Austausch liegt an Position 131 vor. Hier handelt es sich um AS-Austausche zwischen Isoleucin (I), Leucin (L) und Valin (V). Alle drei sind unpolare, aliphatische (kettenförmige) AS und besitzen eine hydrophoben Rest. Weitere isofunktionelle Austausche tauchen zwischen Arginin (R) – Lysin (K), beide basisch, Phenylalanin (F) – Tryptophan (W), beide aromatisch und Isoleucin – Leucin – Valin – Alanin (A), alle vier unpolar aromatisch, auf.

Ramazzina at all vermuten drei an der Katalyse von OHCU zu (S)-Allantoin beteiligten AS, welche im aktiven Zentrum der OHCU Decarboxylase sitzen. Bei diesen drei Aminosäuren handelt es sich um Histidin-67 (H-67), Glutaminsäure-87 (E-87) und Arginin-161 (R-161), wobei sich die Positionsangaben der AS auf ein Alignment aus den Materialien von Ramazzina at all beziehen. Doch anhand dieser Positionen können sich diese drei AS in den MSAs aufspüren lassen. Demnach kann man vermuten, dass es sich im MSA um H-62, E-82 und R-157 handelt und die 100%ige Konservierung der AS E-82 und R-157 dies bestätigen würde. Doch H ist nicht 100%ig konserviert und die zweite sowie die neunzehnte Sequenz besitzen an dieser Position kein H. Des Weiteren taucht ein Y (Tyrosin) in der achtzehnten Sequenz auf.

Eukaryonten. Fungi.

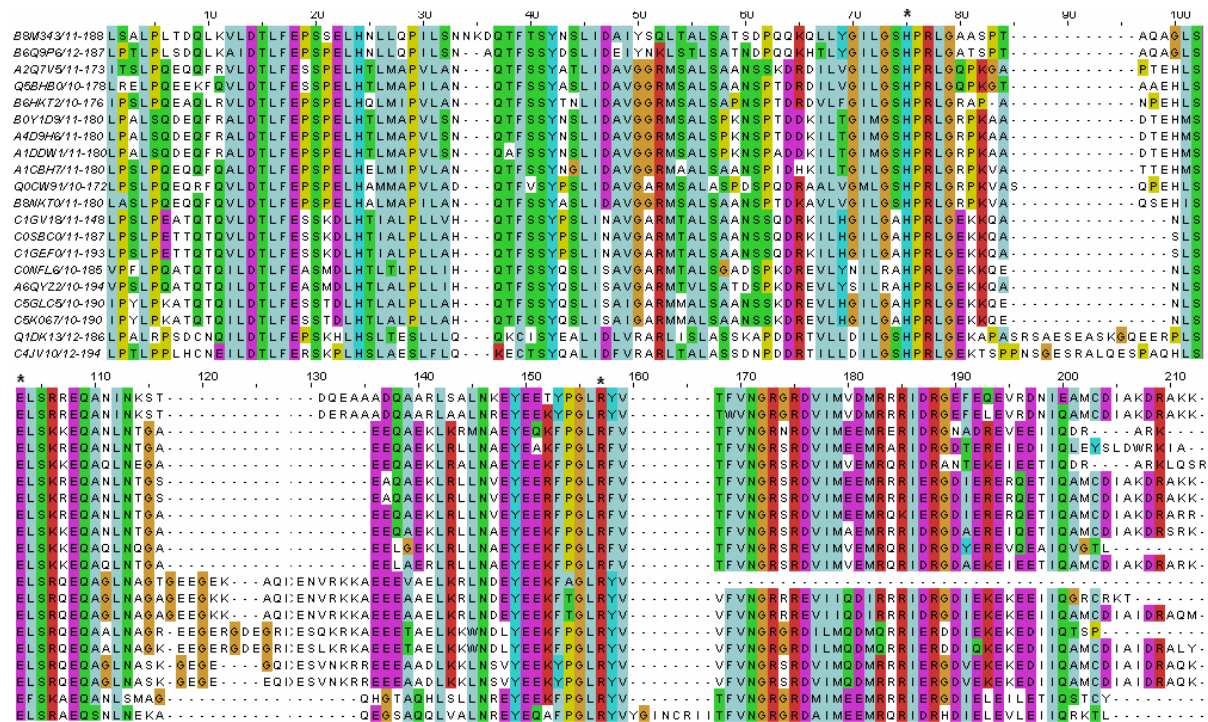


Abbildung 30 JalView Ansicht des *Fungi*-OHCDec-MSA, welches mittels ClustalW erstellt wurde. Mit schwarzem Stern markierte AS sind laut [16] vermutlich an Katalyse beteiligt [23].

Das MSA (siehe Abbildung 30) der Untergruppe *Fungi* umfasst 20 Sequenzen mit einer durchschnittlichen Gesamtlänge von 171 AS. Insgesamt wurden 32 zu 100% konservierte AS gefunden (siehe Tabelle 5). Dazu zählen die fünf aufeinander folgenden AS Histidin (H), Prolin (P), Arginin (R), Leucin (L) und Glycin (G), an den Positionen 75 bis 79 im Alignment. Auffällig ist, dass die Konservierung meist als Blöcke auftreten (siehe Abbildung 24). *Beispiel:* An Position 14 bis 18 ist die Konservierungsanzahl durchgängig, Position 41 bis 45 enthält vier Konservierungen, Position 75 bis 79 (siehe oben), und Position 102 bis 109 enthält 5 100%-ige Konservierungen. Weiterhin gibt es isofunktionelle Austausche unter anderem an den Positionen 1, 12 und 30. Hier liegt der AS Austausch zwischen Isoleucin, Valin und Leucin vor. Besonders dieser AS-Austausch tauchte auch mehrfach in dem MSA der Metazoa, und im weiteren Verlauf auch in den Bakterien-MSA, auf. Folglich ähneln sich die beiden Gruppe der Eukaryonten in dem Punkt der isofunktionellen Austausche. Die vermutlich an der Katalyse beteiligten AS könnten H-75, E-103 und R-157 sein.

Prokaryonten. Actinobacteria.

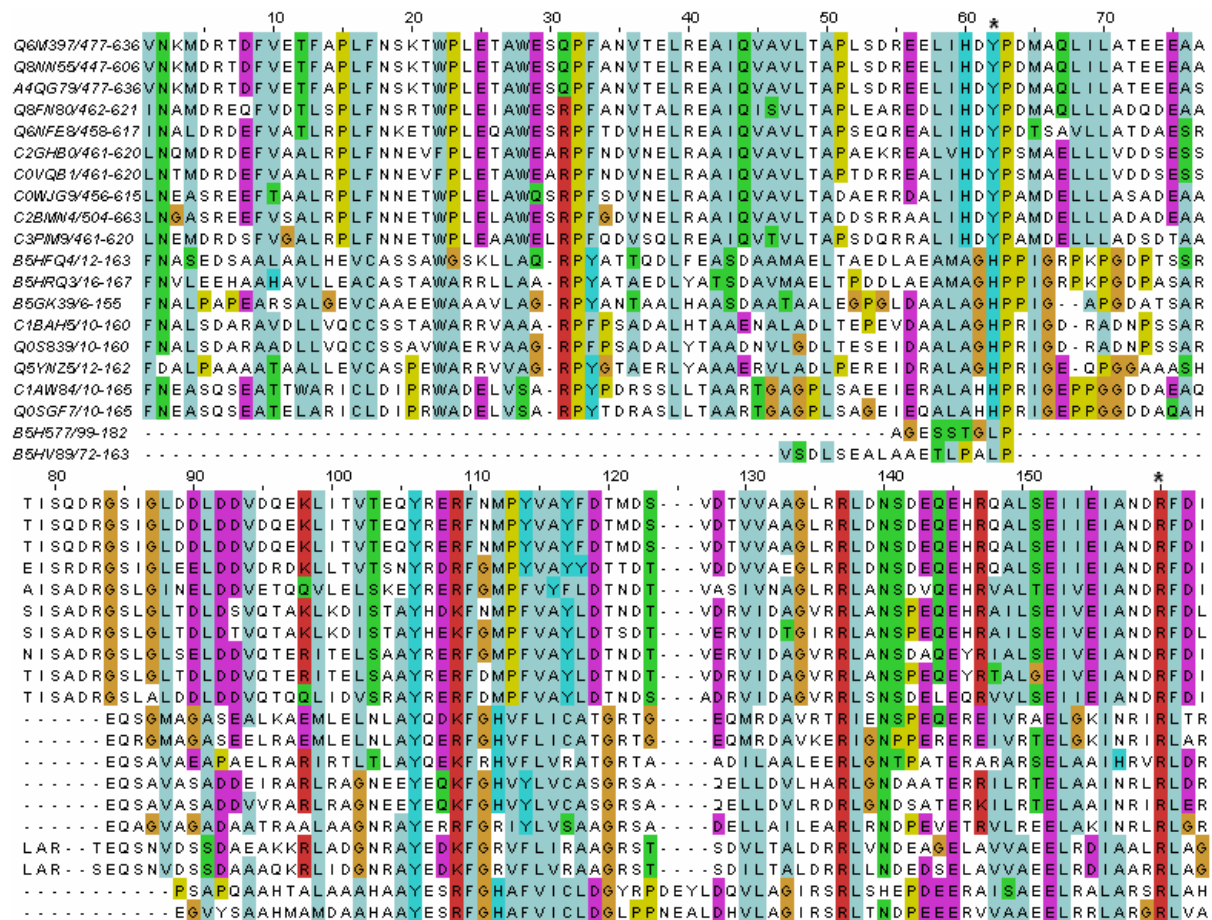


Abbildung 31 JalView Ansicht des *Actinobacteria*-OHCU-Dec-MSA, welches mittels ClustalW erstellt wurde. Mit schwarzem Stern markierte AS sind vermutlich [16] an Katalyse beteiligt [23].

Die 20 AS-Sequenzen des *Actinobacteria* MSA (siehe Abbildung 31) zeigen im Schnitt 148 AS lange Sequenzen auf. Es existieren 7 zu 100% konservierte AS (siehe Tabelle 5), wie zum Beispiel an den Positionen 63 (Prolin) und 106 (Tyrosin). Im Vergleich zu *Metazoa* (18) und *Fungi* (32) treten beträchtlich wenig 100%-ige Konservierungen auf. Isofunktionelle Austausche bestehen unter anderem an den Positionen 109 zwischen Lysin - Arginin und 138 zwischen Valin – Leucin. Die laut Ramazzina at all vermutlich an der Katalyse von OHCU zu (S)-Allantoin beteiligten AS Histidin (H) und Glutaminsäure (E) sind nur zu erahnen. Die AS Arginin (R) ist hingegen 100%ig konserviert und daher gut auffindbar.

Prokaryonten. Proteobacteria.

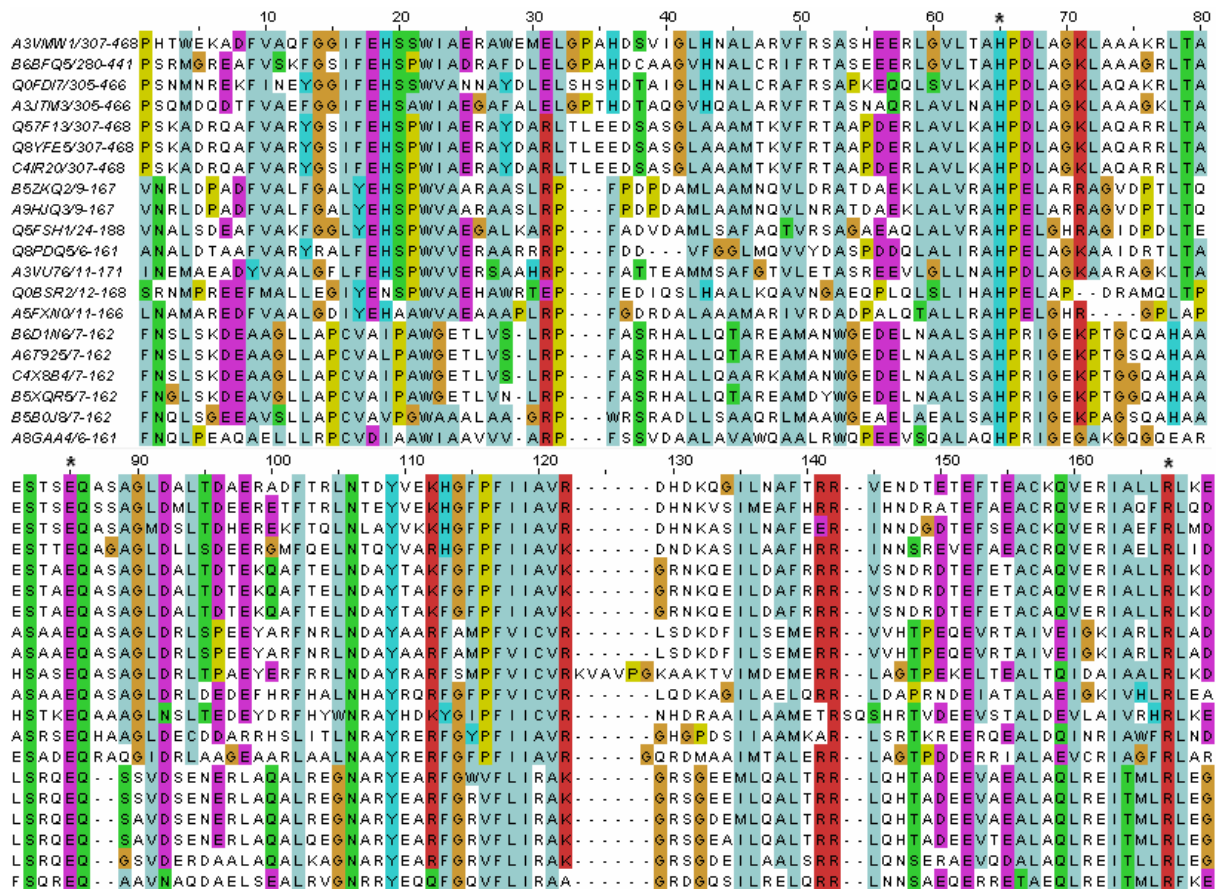


Abbildung 32 JalView Ansicht des *Proteobacteria* MSA, welches mittels ClustalW erstellt wurde. Mit schwarzem Stern markierte AS sind vermutlich [16] an Katalyse beteiligt [23].

Die durchschnittliche Sequenzlänge der 20 AS-Sequenzen des MSA *Proteobacteria* (siehe Abbildung 32) beträgt 158 AS. Das MSA weist 14 100%-ig konservierte AS auf, darunter zählt an Position 141 Arginin und an Position 152 Glutaminsäure. Weiterhin gibt es isofunktionelle Austausche zwischen den AS an Position 62 zwischen Leucin – Valin – Isoleucin sowie an Position 157 zwischen Alanin – Valin – Leucin – Isoleucin. Die vermeintlich an der Katalyse teilnehmenden AS können H-65, E-85 und R-167 lauten.

Zur genaueren Aussage über die drei an der Katalyse teilhabenden Aminosäuren und Aminosäuren, welche der Stabilität des Enzyms beitragen, wurde eine Tabelle (siehe Tabelle 5) der 100%igen Konservierungen sowie ein WebLOGO (siehe Abbildung 33), basierend auf dem MSA aller 80 Sequenzen, erstellt.

Tabelle 5 Tabellarische Darstellung der 100%-ig konservierten AS von *Metazoa*, *Fungi*, *Actinobacteria*, *Proteobacteria*.

<i>Metazoa</i>	<i>Fungi</i>	<i>Actinobacteria</i>	<i>Proteobacteria</i>
L-65	D-14	P-63	W-22
L-75	T-15	Y-106	H-65
S-79	L-16	F-110	P-66
E-82	F-17	R-137	S-82
G-87	E-18	E-145	E-85
L-91	S-20	E-152	Q-86
Y-101	L-23	R-160	N-106
F-110	H-24		Y-109
P-113	S-41		F-117
F-114	Y-42		I-119
V-115	L-45		R-142
R-134	I-46		E-152
E-142	L-56		I-163
E-149	L-68		R-167
K-152	H-75		
I-153	P-76		
L-156	R-77		
R-157	L-78		
	G-79		
	S-102		
	E-103		
	S-105		
	E-108		
	Q-109		
	L-142		
	N-146		
	Y-149		
	E-150		
	G-155		
	L-156		
	R-157		
Farbgebung:			
übereinstimmende AS zwischen M. & F.			
übereinstimmende AS zwischen M. & F. & P.			
übereinstimmende AS zwischen M., F., P. & A.			
übereinstimmende AS zwischen M., A. & P.			
übereinstimmende AS zwischen A. & P.			

Wie der Tabelle 5 entnommen werden kann, treten im Vergleich *Metazoa* und *Fungi* drei Leucin Konservierungen auf. Hierbei handelt es sich um die drei grün gefärbten Felder. Aus der Betrachtung der Konservierungen der *Actinobacteria* und *Proteobacteria* zeigt sich, dass es in diesen beiden Gruppen keine Konservierung der Aminosäure Leucin existiert. Folglich könnte es sich bei Leucin um eine Konservierung handeln, welche nur in den Eukaryonten auftritt. So könnte man einzig und allein die eukaryontischen Sequenzen von den prokaryontischen Sequenzen über die Konservierung des Leucins unterscheiden.

Aus dem Vergleich von *Actinobacteria* und *Proteobacteria* ergibt sich eine übereinstimmende Konservierung, welche nicht bei den Eukaryonten auftritt. Diese Konservierung entspricht den gelb unterlegten Feldern von Prolin (P).

Die orange unterlegten Felder von Serin (S) und Glutaminsäure (E) zeigen eine Konservierung die nur die *Metazoa*, *Fungi* und *Proteobacteria* betrifft, da keine 100%ige Konservierung dieser Aminosäuren bei der Gruppe *Actinobacteria* zu finden war. Die blau unterlegten Felder betreffen die Aminosäure Phenylalanin (F). Diese Aminosäure ist, außer in der Gruppe der *Fungi*, hochkonserviert und tritt nach der ebenfalls hochkonservierten Aminosäure Tyrosin (Y) auf. Die Aminosäure F existiert zwar in den *Fungi* Sequenzen (Position 153); sie ist jedoch nur zu 80% konserviert, doch wird sie in manchen Sequenzen von der Aminosäure Y ausgetauscht. Hierbei handelt es sich allerdings um einen isofunktionellen Austausch, da beide Aminosäuren Aromaten sind und hydrophobe Reste besitzen. Dieser Austausch beeinflusst folglich in keiner Weise die Funktionalität des Enzyms. Somit ist zwar F nicht 100%ig konserviert, doch trägt dieser Austausch zum Funktionserhalt des Enzyms bei. Die rosa unterlegten Felder betreffen die Aminosäuren Tyrosin (Y) und Arginin (R). Diese beiden Aminosäuren treten in allen Sequenzen auf. Die letzte 100%ig konservierte Aminosäure einer jeden Sequenz stellt R dar. Aus den Papern von Ramazzina at all und Kim at all (siehe 3.1) ging hervor, dass vermutlich die Aminosäure Histidin (H), Glutaminsäure (E) und Arginin (R) die an der Katalyse beteiligten Aminosäuren sind.

Insgesamt betrachtet werden bei der Gruppe *Fungi* die meisten 100%ig konservierten Aminosäure gezählt. Daraus lässt sich vermuten, dass diese Gruppe die eventuell wenigsten Mutationen im Laufe der Evolution durchlebt hat bzw. wurden die funktionsbeeinträchtigten Mutationen durch natürliche Selektion behoben.

Die Aminosäure H-75, zu 81% konserviert, wird in manchen Sequenzen durch Y ausgetauscht, vor allem in den *Actinobacteria* Sequenzen liegt dieser Austausch vor. Beide Aminosäuren hingegen zeigen in ihrer Physikochemie auf den ersten Blick eher unterschiedliche Eigenschaften. Beispielsweise verhält sich H stark hydrophil, Y

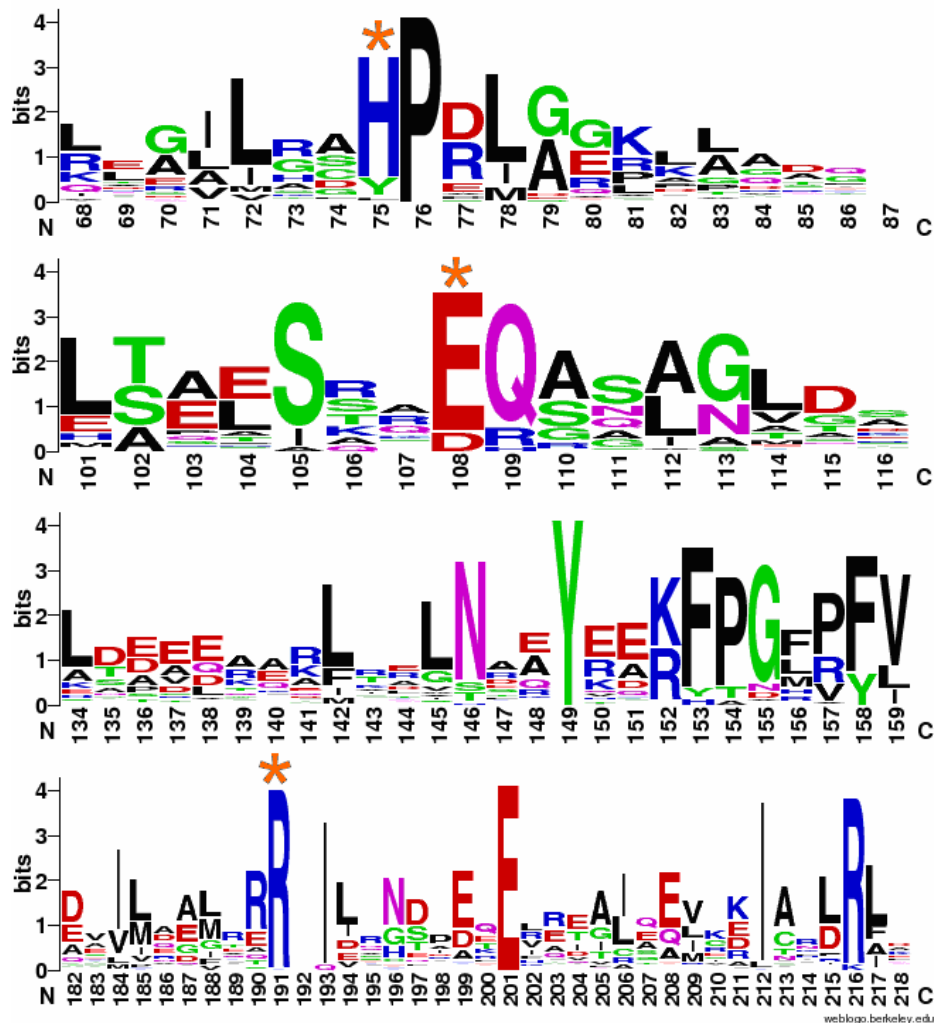


Abbildung 33 WebLogo (zwischen Position 68 und 218) basierend auf einem MSA der vier Gruppen. Je größer ein Buchstabe an einer Position, desto häufiger tritt diese AS an der jeweiligen Position einer Sequenz auf und desto höher ist ihre Konservierung, die automatische Farbgebung entspricht der Hydrophobizität der AS, aus dem MSA stammende Lücken bzw. –bereiche tauchen im Logo nicht auf [24]. Orangefarbener Stern kennzeichnet an Katalyse vermutete AS.

Die Aminosäure H-75, zu 81% konserviert, wird in manchen Sequenzen durch Y ausgetauscht, vor allem in den *Actinobacteria* Sequenzen liegt dieser Austausch vor. Beide Aminosäuren hingegen zeigen in ihrer Physikochemie auf den ersten Blick eher unterschiedliche Eigenschaften. Beispielsweise verhält sich H stark hydrophil, Y

hingegen eher hydrophob. Folglich müsste bzw. würde die katalytische Funktion des Enzyms beeinträchtigt werden. Doch eventuell wird dies verhindert. Histidin (H) weist einen leicht protonierbaren Imidazol-Stickstoff auf. Tyrosin (Y) besitzt einen hydrophoben Rest mit einer phenolischen Hydroxylgruppe. Dieser Rest kann ebenso protoniert werden, möglicherweise aber gegenüber dem H beschwerlicher, sodass dessen ungeachtet die katalytische Eigenschaft des Enzyms dennoch nicht gestört wird. Ferner besitzen beide AS eine aromatische Struktur und stimmen in ihrer Polarität, beide polar, überein [4].

Die zu 85% konservierte AS E-108 wird, vor allem in den *Actinobacteria* Sequenzen, mit der AS D getauscht. Da sich beide AS jedoch stark in ihren physikochemischen Eigenschaften ähneln, gilt dieser Austausch, der weder auf die Struktur des Enzyms noch auf die Funktion einen negativen Einfluss übt, als isofunktionell.

Weiterhin besteht neben der 97%-igen Konservierung der ebenfalls an der Katalyse teilnehmenden AS R-191 eine 100%-ige Konservierung der AS Y-149. Diese AS trägt voraussichtlich einen sehr starken Anteil zur Stabilisierung der Struktur der OHCU Decarboxylase teil. Ebenso der Stabilisierung beitragend sind die hochkonservierten AS P-76 (98% konserviert) und E-201 (98% konserviert). Auch diese hochkonservierten AS wurden ebenso von Ramazzina et al. vermutet und können so als Beweis unterstützend der Annahmen von Ramazzina dienen.

4.2 Erstellung eines phylogenetischen Baumes auf der Ebene der Sequenzen

Aufgrund der in Punkt 4.1 vorgestellten Alignments, der Tabelle 5 und des WebLogos, welche den Grad der Konservierungen einiger (bedeutender) Aminosäuren zeigen, erhielt man einen ersten Einblick in den Sequenzaufbau der Familie der OHCU Decarboxylasen. Um nun einen evolutionären Hintergrund zu gewinnen, wurde ein phylogenetischer Baum auf Grundlage des Neighbor Joinings (NJ, siehe Punkt 1.2.3), entwickelt (siehe Abbildung 34).



Abbildung 34 Stark verkleinerter Phylogenetischer NJ Baum (siehe Anhang-größere Abbildung 47) der Familie der OHCU Decarboxylasen, basierend auf dem MSA der 80 ausgewählten Sequenzen [19]. Namensgebung der Gruppierungen: P1, P2 = *Proteobacteria*, A1, A2 = *Actinobacteria*, F = *Fungi*, M = *Metazoa*; Pfeile: rot = gemeinsamer Knoten von A/F/M/P, orange = gemeinsamer Knoten von P1 und A1, rosa = Knoten von F, blau = Knoten von M, grau = Divergenzzeit; grüne Box zeigt Ausnahme-Sequenz von A1.

Aus der Abbildung des phylogenetischen Baums geht hervor, dass alle 80 Sequenzen der insgesamt vier Gruppen *Metazoa*, *Fungi*, *Actinobacteria* und *Proteobacteria* auf einen gemeinsamen Vorfahren zurückgehen, da sie aus einem

Knoten (roter Pfeil) hervorgehen. Weiterhin lassen sich die 20 Sequenzen der *Metazoa* (blaues M) und die 20 Sequenzen der *Fungi* (rosa F) als jeweils eine Gruppe im Baum erfassen, d.h. sowohl die Fungi Sequenzen als auch die Metazoa Sequenzen gehen jeweils aus einem Knoten (Pfeil rosa bzw. blau) hervor und teilen diesen Knoten mit keiner anderen Gruppe oder Mitgliedern einer anderen Gruppe. Aus der Gruppe der *Actinobacteria* sowie der *Proteobacteria* kristallisieren sich je zwei Untergruppen heraus: P1 & P2, A1 & A2. Festzustellen ist, dass die Gruppen P1 und A1 sehr nahe verwandt sein müssen, da sie aus einem gemeinsamen Vorläufer hervorgehen, dies entspricht der Monophylie, indem sie sich einen Knoten (markiert mittels orangefarbenen Pfeil) teilen, mit Ausnahme von B5H588 und B5HV89 (grün gerahmt). B5HV89 teilt sich einen gemeinsamen Knoten mit der Gruppe F. Folglich verhält sich die Sequenz B5HV89 zur Gruppe F monophyletisch, d.h. sie besitzen den gleichen Vorfahren. Ebenso monophyletisch zueinander verhalten sich P1/A1 zu B5HV89/F. Ferner liegt eine Paraphylie zwischen der Sequenz B5HV89 und A1 vor. Eine Paraphylie bedeutet, dass sich die Mitglieder der Gruppe A1 (inklusive B5HV89) auf eine letzte gemeinsame Stammform zurückführen lassen, welche sie jedoch mit Angehörigen einer anderen Gruppierung, in diesem Falle P1, teilen. Aus dem Vergleich mit dem MSA (siehe Abbildung 31) der Gruppe *Actinobacteria* resultiert, dass die zwei Sequenzen B5H577 und B5HV89, welche sich polyphyletisch zueinander verhalten, die kürzesten Sequenzen der *Actinobacteria* sind. An Position 62 befindet sich bei diesen beiden Sequenzen ein Leucin (L), alle anderen Sequenzen der *Actinobacteria* Gruppe besitzen hier ein Histidin (H) bzw. Tyrosin (Y). H wurde von Ramazzina at all als mutmaßliche AS in der Katalyse der OHCU Decarboxylase definiert (siehe vorherigen Punkt und Punkte 3.1 & 3.2). H taucht bei B5H577 und B5HV89 im MSA erst an Position 112 auf. Darüber hinaus ist festzustellen, dass der Gruppe A2 die ersten zehn Sequenzen (siehe MSA, Abbildung 31) zugehören. Diese besitzen im Gegensatz zu den Sequenzen 11-18 (Nummerierung laut Reihenfolge des MSA in Abbildung 31) an Position 62 (AS vermutlich an Katalyse beteiligt) ein Tyrosin, anstatt eines Histidins. Weitere „spezielle“ AS-Unterschiede bestehen unter anderen an den Positionen 15 (P), 25 (E), 44 (Q) und 60 (H). Eventuell ist H-60 an der Katalyse beteiligt, anstelle eines AS-Austauschs an der Position 62 von H durch Y. Da H sehr viel leichter protoniert werden kann als Y und so die Annahme von Ramazzina at all unterstützen würde. Auch im MSA Vergleich der Gruppe P1 (Sequenz eins bis 14, Ausnahme 13 -

Q0BSR2 - Outgroup) und P2 (Sequenz 15 bis 20) der *Proteobacteria*, welche die Outgroup des Stammbaumes der OHCU Decarboxylase Familie stellt, gibt es erkennbare AS - Unterschiede bzw. - Austausche an einzelnen Positionen. Unterschiede lokalisieren sich zum Beispiel an Position 19, hier besitzen P2 die AS H, die P1 hingegen weisen Isoleucin(I) / L / Valin (V) auf oder an der Position 105. Hier tritt bei P1 die AS Glycin (G) auf, in Gruppe P2 handelt es sich jedoch um L oder Tryptophan (W).

Die Outgroup des Stammbaumes, d.h. die Außengruppe, stellt die *Proteobacteria* Sequenz Q0BSR2 dar. Diese Außengruppe soll sich von einem früheren gemeinsamen Vorfahren mit der Innengruppe (alle anderen Sequenzen) ableiten lassen, dennoch ist die Außengruppe nicht nah mit der Innengruppe verwandt und entspricht dem phylogenetisch ältesten Ast des Baumes. Eine Außengruppe entsteht mit dem Rooting des Baumes und verleiht diesem eine Richtung sowie historische Bedeutung [14]. Weiterhin spaltet sich die *Actinobacteria* Sequenz A5FXN0 von den restlichen 12 Sequenzen der Gruppe A2 ab.

5. Betrachtung der Familie OHCu Decarboxylase auf Ebene der Energieprofile

5.1 Erstellung der Energieprofile basierend auf den Strukturen 2o70 und 2q37

Alle nach folgenden Energieprofile (EPs) wurden mithilfe des Tools eCalc auf dem ePros Server erstellt. Laut der Pfam Datenbank existieren 19 Strukturen. Dabei handelt es sich im Grunde jedoch mehr oder weniger um vier Strukturen: 2o70, 2o73, 2o74 und 2q37, da jede Kette von 2o70, 2o73 und 2o74 als eigene Struktur angesehen wird, sie jedoch keine Unterschiede zueinander aufweisen. Die Strukturen 2o70, 2o73 und 2o74 sind ein und dieselbe Struktur, welche von der Zebrafischsequenz der OHCu Decarboxylase abgeleitet wurden. Der Unterschied zwischen den drei Strukturen besteht lediglich darin, dass 2o73 in Anwesenheit von Allantoin und 2o74 in Gegenwart von Guanin dargestellt werden. Daher werden in den folgenden Untersuchungen, bezüglich der Energieprofile, nur die Strukturen 2o70 und 2q37 berücksichtigt.

Als Darstellungsform der Energieprofile wurde die Quantil-Färbung zur Quantifizierung der Energie gewählt. Hierbei wurde das Energiespektrum in vier Quantile unterteilt. Das hochenergetischste erste Quantil (Q1), wird rot, das zweite Quantil (Q2) wird grün, dritte Quantil (Q3) wird gelb und das niederenergetischste vierte Quantil (Q4) wird blau koloriert.

Laut einem Theorem scheint die Energie äquivalent zur Stabilität einer AS zu sein, mit welcher diese AS in der Struktur verankert ist. Dabei besitzen fest in die Struktur integrierte AS bzw. ganze Sequenzabschnitte eine niedrigere Energie. Zu solchen Strukturen zählen vor allem α -Helices und β -Faltblätter, da diese regelmäßig in ihrer Struktur sind und durch zusätzliche Wasserstoffbrückenbindungen an Stabilität gewinnen und daher auch stabiler sind als Coils [5].

Die von der Zebrafisch OHCu Decarboxylase ermittelte Struktur 2o70 ist in der Abbildung 35 als Energieprofil über die gesamte Sequenz zu sehen. Die drei schwarz markierten AS sind jene, welche vermutlich eine große katalytische Rolle in der Decarboxylierung des OHCu zu (S)-Allantoin spielen. Dabei handelt es sich um H-67, E-87 und R-161.

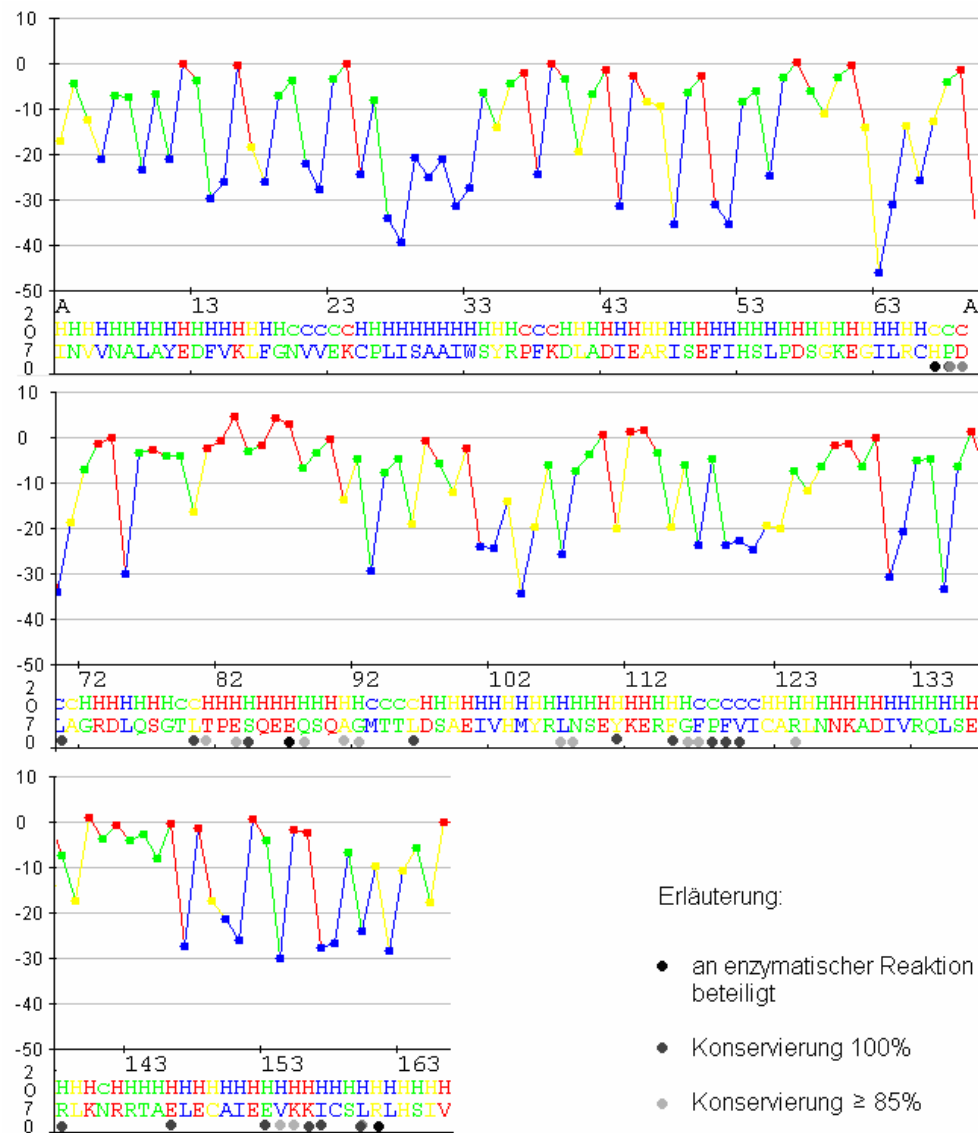


Abbildung 35 EP der Struktur 2o70 (PDB-ID) - Energie (-50 bis 10) gegen AS-Sequenz (untere Zeile) mit zugehöriger Sekundärstruktur (obere Zeile). Farbgebung (hoch- zu niederenergetisch): Q1=rot, Q2=grün, Q3=gelb, Q4=blau [25]; schwarzer Punkt = vermutlich an Katalyse beteiligte AS [It. Ramazzina at all], dunkelgrauer Punkt = 100%-ige Konservierung, hellgrauer Punkt = hoher Konservierung.

Betrachtet man die Energien der drei AS, so entsprechen H-67 und R-161 dem niederenergetischen Q3 und E-87 dem hochenergetischen Q1. H-67 befindet sich laut dem EP in einem Coil zwischen der dritten und vierten Helix (Sekundärstruktur, siehe auch Punkt 3.1). E-87 und R-161 werden jeweils in einer Helix lokalisiert. Nach Annahmen von Ramazzina nimmt H-67 an der enzymatischen Reaktion teil, da der AS-Rest von H protoniert wird und H-67 somit ein Proton zum Enol-Zwischenprodukt liefert (ausführliche Reaktion, siehe Punkt 3.2). Da H-67 nicht in einer stabilen Helix-

Struktur eingebettet liegt, sondern in einer instabileren Coil-Struktur lokalisiert wird, kann diese AS womöglich als Protonendonor in der enzymatischen Reaktion agieren.

Betrachtet man die 100%-ig konservierten und die hochkonservierten AS (Konservierungsgrad $\geq 85\%$) in Bezug auf deren energetische Verteilung in den einzelnen Quantilen, kann man spekulativ behaupten, dass weniger hochkonservierte AS in das hochenergetische Q1 liegen als in den niederenergetischen Quantilen. Insgesamt betrachtet lassen sich jedoch leider keine eindeutigen Aussagen auf das Energiequantil-Konservierungsverhalten treffen (siehe Tabelle 6).

Tabelle 6 Verteilung zwischen Energiequantile und Konservierungsgrad, Q1 = ↑Energie bis Q4 = ↓Energie.

	Konservierungsgrad			
Quantil	85%	90%	95%	100%
Q1	1	1	2	3
Q2	2	-	4	5
Q3	2	2	-	5
Q4	1	1	1	5

Des Weiteren werden in diesem Energieprofil von 2o70 nur acht Helices ersichtlich, laut Ramazzina handelt es sich jedoch um zehn α -Helices und eine 3.10-Helix. Es ist bislang leider noch nicht möglich, genauere Aussagen im Bezug auf die Sekundärstrukturelemente mit Visualisierungsprogrammen in Zusammenhang mit EPs zumachen, denn beispielsweise können manche Coil-Strukturen wie helicale Strukturen aussehen, was die Zuordnung zum eigentlichen Sekundärstrukturelement erschwert und auch eigentliche helicale Strukturen können als Coils erkannt werden.

Die Abbildung 36 zeigt das Energieprofil zur Struktur 2q37, der *Arabidopsis thaliana* OHCU Decarboxylase. Im Bereich von 58 bis 81 fehlt ein EP- bzw. Sequenzbereich, da an dieser Stelle keine genaueren Daten zur Verfügung standen. Die an der Katalyse vermutlich teilnehmenden AS Histidin (H), Glutaminsäure (E) und Arginin (R) sind, wie im EP von 2o70, mit einem schwarzen Punkt gekennzeichnet. Im Falle der Struktur 2q37 wird H an Position 58, E an Position 80 und R an Position 153 lokalisiert.

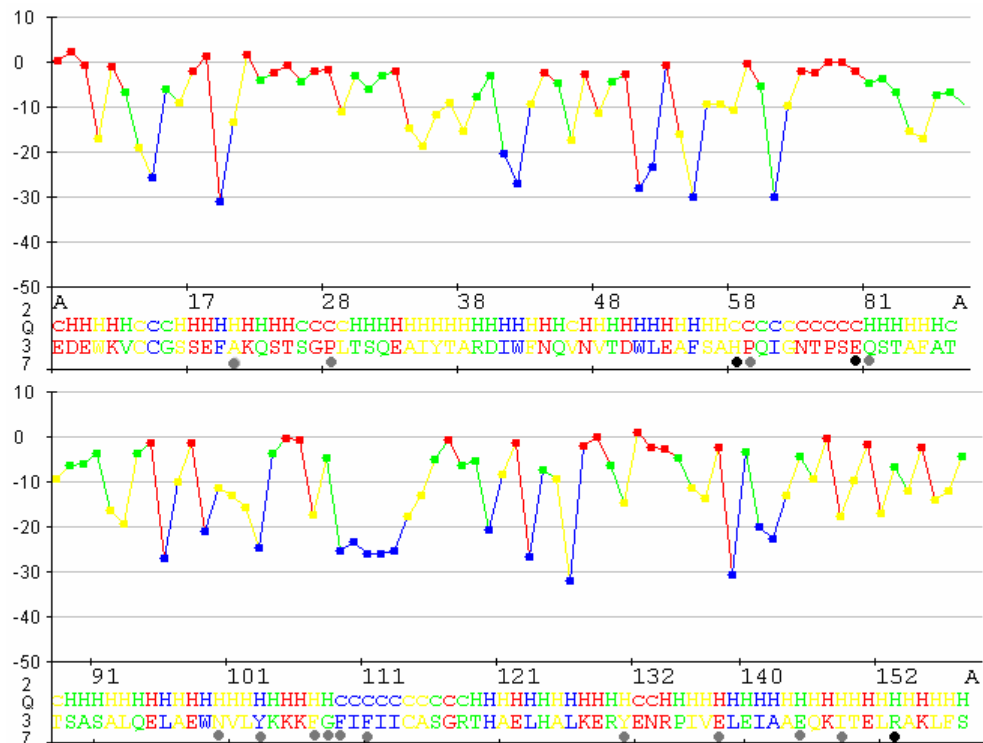


Abbildung 36 EP der Struktur 2q37 (PDB-ID) - Energie (-50 bis 10) gegen AS-Sequenz. Farbgebung (hoch- zu niederenergetisch): **Q1=rot**, **Q2=grün**, **Q3=gelb**, **Q4=blau** [25]; schwarzer Punkt = vermutlich an Katalyse beteiligte AS [lt. Ramazzina at all], dunkelgrauer Punkt = 100%-ige Konservierung [lt. Kim at all].

Im Vergleich zur Struktur 2o70 ergibt sich, dass H-58 ebenfalls in einem Coil und R-153 in einer Helix eingebettet vorliegen. Die AS E-80 befindet sich im Gegensatz zu E-87 (2o70), welche in einer Helix liegt, jedoch in einem Coil. Stellt man die Energien (siehe Tabelle7) der AS gegenüber, so fällt H-58 in das Q3, E-80 in das Q1 und R-153 in das Q2.

Tabelle 7 Tabellarische Gegenüberstellung der Strukturen 2o70 und 2q37 in Bezug auf der energetischen Quantil-Verteilung der drei an der Katalyse vermuteten AS H, E, R.

AS	2o70	2q37
H-67 / 58	Q3	Q3
E-87 / 80	Q1	Q1
R-161 / 153	Q3	Q2

Der Vergleich zeigt, dass H in beiden Fällen eher niederenergetisch ist, da H in Q3 liegt. Weiterhin gibt es eine Übereinstimmung der Energiequantile von E-87 und E-80, beide befinden sich im hochenergetischen Q1. Ein Unterschied besteht bei R, denn R-161 (Q3) besitzt eine niedrigere Energie als R-153 (Q2). Schaut man in die EPs erkennt man aber, dass der energetische Unterschied eher gering ausfällt, da R-161 bei etwa -10 und R-153 bei ca. -7 liegt.

Tabelle 8 Einordnung aller AS der Strukturen 2o70 und 2q37 in die entsprechenden Sekundärstrukturelemente (SSE) und Energiequantile zur Erfassung der Energienverteilung in den SSE.

Struktur	SSE	Q1	Q2	Q3	Q4
2o70	Helix	31	41	24	38
	Coil	4	7	3	9
2q37	Helix	25	27	35	13
	Coil	12	5	6	7

Die Tabelle 8 zeigt eine Verteilung der AS-Energien bezüglich der SSE in den einzelnen Quantilen, wobei Q1 das hochenergetischste und Q4 das niederenergetischste Quantil darstellen. Wie anfangs erwähnt, sind Coil-Strukturen eher hochenergetisch, da sie aufgrund fehlender zusätzlicher Wasserstoffbrückenbindungen instabiler sind als Helices. Bei Struktur 2q37 trifft diese Vermutung zu, bei Struktur 2o70 eher nicht. Vermutlich werden in dem EP von 2o70 manche helicale Strukturen als Coils erkannt. Dies könnte eine Erklärung dafür sein, dass nur acht Helices im EP entdeckt wurden und nicht zehn. Folglich könnte es sich bei den niederenergetischen, als Coil detektierten Bereichen eher um helicale Strukturen handeln, da eine Helix, laut dem eingangs erwähnten Theorem, eine niedrigere Energie aufweisen müsste als eine Coil-Struktur. Betrachtet man nun die Aufteilung der AS-Energien der einzelnen Quantile in der Helix-Konformation, so lässt sich leider nicht genau sagen, ob es sich bei helicalen Strukturen eindeutig um niederenergetische Bereiche handelt oder nicht, denn dafür sind die AS sehr gestreut verteilt.

5.2 Erstellung eines Multiplen Energieprofil Alignments basierend auf Energieprofilen

Paarweises Energieprofilalignment basierend auf den Energieprofilen der Strukturen 2o70 und 2q37. Die in den nachfolgenden Abbildungen dargestellten Energieprofil Alignments wurden mittels des Tools eAlign des ePros-Servers angefertigt, wobei die Struktur bzw. Sequenz - Eingabedaten der PDB (speziell die PDB ID) entnommen wurden.

Die Abbildung 37 zeigt ein paarweises Energieprofilalignment (PEPAL) der Strukturen 2o70 und 2o74. Beide Strukturen sind sich sehr ähnlich, denn es handelt sich hier um zwei identische Sequenzen, da die Struktur 2o74 ebenso wie Struktur 2o70 eine Zebrafisch-Sequenz ist, mit dem Unterschied, dass in der dreidimensionalen Darstellung von 2o74 das Purin Guanin (Ausgangstoff des Purinkatabolismus) anwesend ist. Die Identität zeigt sich in der Deckungsgleichheit der Energiekurven und in der Übereinstimmung der Sequenzen, sowie der SSE.

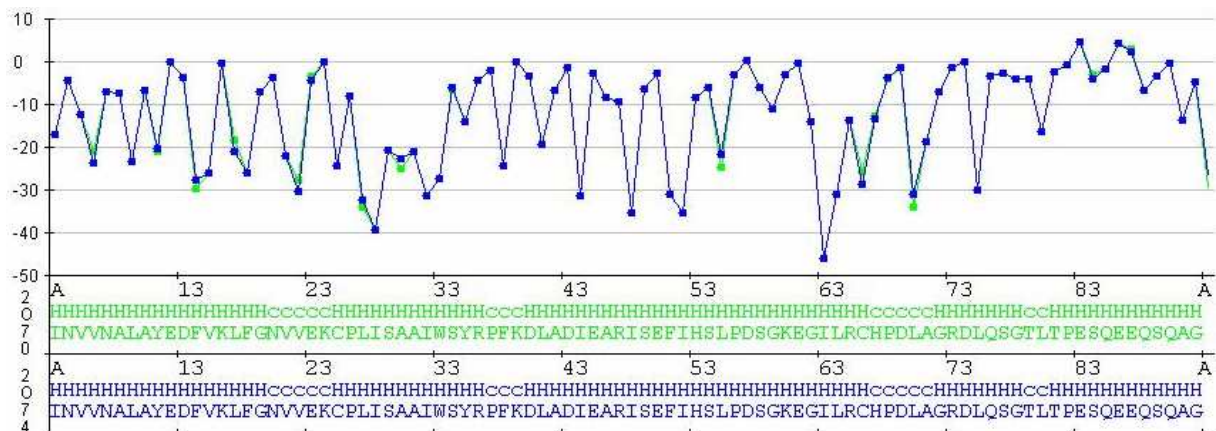


Abbildung 37 Ausschnitt des PEPAL der zwei Strukturen 2o70 (grün) und 2o74 (blau). Darstellung des EP-Alignments: Diagramm Energie (-50 bis 10) gegen AS-Sequenzen von Struktur 2o70 (grün) & 2o74 (blau). Über der AS-Sequenz befindet sich die Abfolge der SSE (H= Helix, C= Coil) [25].

Im PEPAL der Strukturen 2o70 und 2q37 (siehe Abbildung 38) wird ersichtlich, dass beide Strukturen Ähnlichkeiten aber auch Unterschiede in ihren Sequenzen und Energieprofilen aufweisen. Die grün dargestellte Sequenz bzw. Energiekurve entspricht der Zebrafisch OHCU Decarboxylasen Struktur 2o70, die blau abgebildete Sequenz bzw. Energiekurve zeigt die OHCU Decarboxylasen Struktur 2q37 von *Arabidopsis thaliana*.

rot umrahmten Bereiche. Diese Bereiche sind in dem Energieprofilalignment in Abbildung 37 die Bereiche, mit der größten Ähnlichkeit (rote Linien). Der beigefarbene Balken (= schwach ähnlich) in Abbildung 37 entspricht zum Teil dem letzten orange gerahmten Bereich in Abbildung 38. Hier besitzen die Energieintervalle der entsprechenden AS einen Abstand von 1. Einen Abstand von > 1 zeigen die gelb gerahmten Positionen in Abbildung 39 an. Betrachtet man die ersten gelb gerahmten AS (Position 2) und vergleicht diese mit dem EP in Abbildung 38, so erkennt man, dass sich die EP-Kurven in diesen Bereichen nicht ähneln.

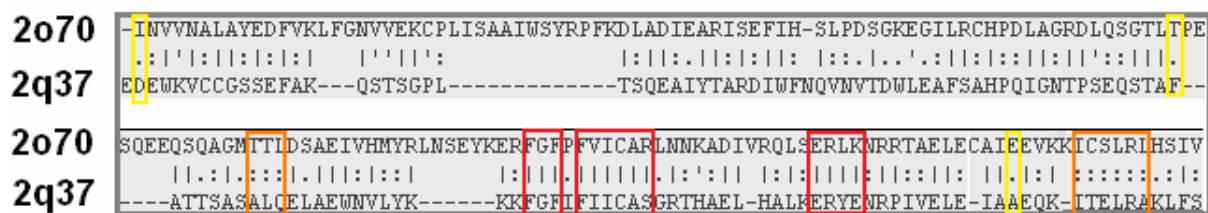


Abbildung 39 Paarweises Alignment von 2o70 und 2q37. „ - “ = Lücke, „ | “ = Intervall beider Energien an dieser Position sind gleich, „ : “ = Abstand beider Intervalle ist 1, „ . “ = Abstand beider Intervalle > 1, „ ’ “ = Intervalle sind entgegengesetzt [25].

Dennoch überwiegen die Bereiche mit großer und mittelgroßer Ähnlichkeit. Dies wird letztendlich auch im ePlot (siehe Abbildung 40), dem Dotplot der EPs, sichtbar. Eine exakt diagonal verlaufende graue Linie entspräche zwei identischen EPs. Im ePlot der Strukturen 2o70 und 2q37 wird eine in etwa diagonal verlaufende Gerade ersichtlich. Dieses ePlot zeigt somit eine relativ große Ähnlichkeit zwischen den EPs beider Strukturen auf. Das wiederum beweist, dass die Sequenzebene von Proteinen nicht immer exakt identisch sein muss, um eine fast übereinstimmende Struktur zu erzielen. Denn wie hier im Beispiel zusehen, zeigen 2o70 und 2q37 sehr wenige Übereinstimmungen in der Sequenzebene. Doch in der EP-Ebene gibt es ähnliche bis sogar sehr ähnliche Bereiche, sodass eine gleiche Funktion der Proteine möglich wird, auch mit unterschiedlichen AS-Abfolgen auf Sequenzebene.

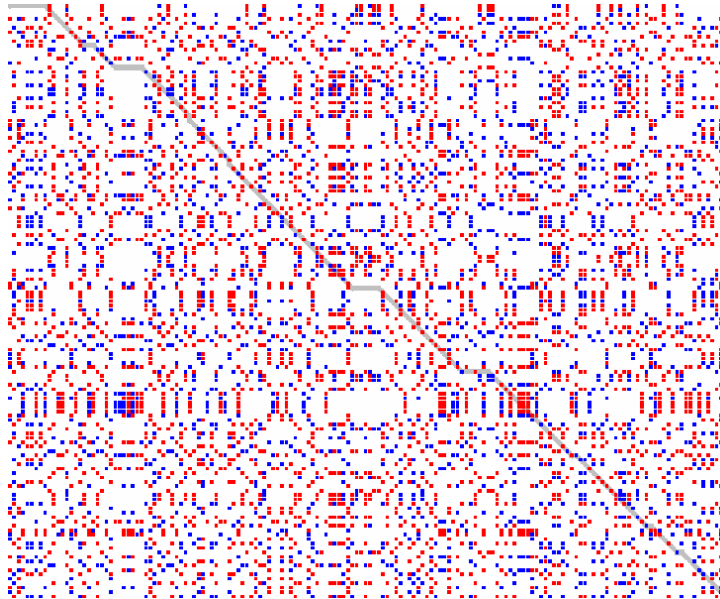


Abbildung 40 ePlot. Das ePlot visualisiert die paarweisen Energie-Scores jeder AS des Proteins 2o70 zu allen AS des Proteins 2q37. **Blaue** bzw. **rote** Punkte kennzeichnen einen paarweisen Energie-Score beider AS von größer **0** bzw. **6**; die grauen Punkte zeigen den Backtracking Pfad [25].

Die Abbildung 41 zeigt die beiden dreidimensionalen Strukturen, beide sind Dimere und verfügen über helicale Strukturen, welche mittel Coils verbunden werden. Im Großen und Ganzen sehen die Strukturen unterschiedlich aus, erfüllen jedoch dieselbe Funktion.

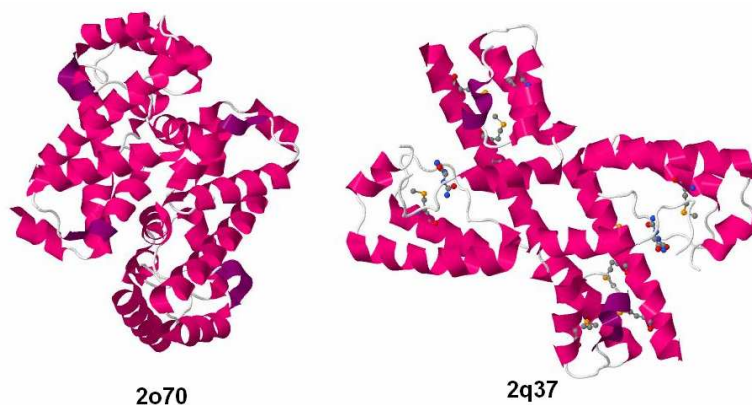


Abbildung 41 3D-Darstellung der Struktur von Zebrafisch- (2o70) und *Arabidopsis thaliana* (2q37) OHCU Decarboxylase [20].

Multiples Energieprofil Alignment.

Für die Erstellung eines multiplen Energieprofil Alignments (MEPAL, siehe Abbildung 42) wurden zunächst die einzelnen 80 EP der 80 Sequenzen kalkuliert, welche anschließend zueinander und zu den zuvor ebenfalls berechneten sechs Energieprofilen der Struktur 2o70 (Ketten A bis F) aligniert wurden. Das entstandene MEPAL ist in der nachfolgenden Abbildung zu sehen.

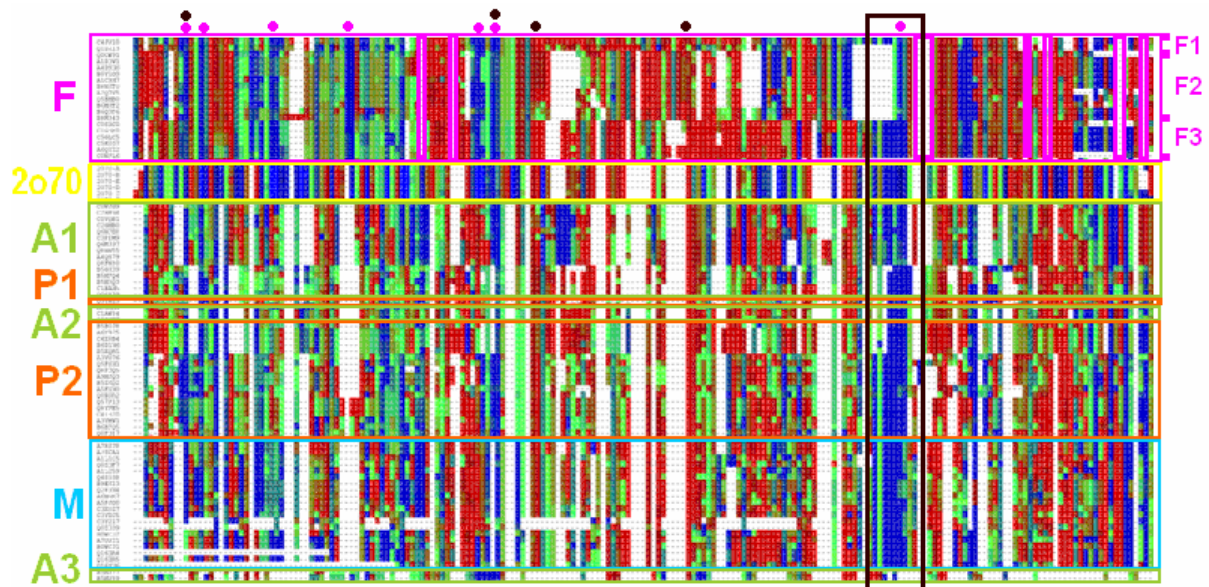


Abbildung 42 Multiples Energieprofil Alignment (MEPAL) basierend auf Sequenz- und Struktur-EPs. Farbgebung: AS = je blauer desto niedriger ist die Energie, je roter desto höher ist die Energie; Rahmenfarben: gelb = Struktur 2o70, rosa = Fungi (F), grün = Actinobacteria (A), blau = Metazoa (M), orange = Proteobacteria (P). Farbgebung: rosa Punkte = Konservierung Fungi, braune Punkte = gleiche Energie in allen Sequenzen, brauner senkrechter Rahmen = konservierter Bereich.

Aus der Abbildung 42 geht hervor, dass alle 20 Sequenzen der Gruppe *Fungi* im MEPAL in einer Gruppe stehen (siehe rosa Rahmen). Sie weisen spezifische Lücken (siehe Abb. 42, senkrechte rosa Rahmen) auf, die in den anderen Organismengruppen so nicht auftreten. Des Weiteren bilden sich, durch eine Art „Blockbildung“, aus der *Fungi* Gruppe zwei bzw. drei Untergruppen heraus. Dabei wird die erste Untergruppe (F1) von Sequenz eins und zwei, die zweite Untergruppe (F2) von Sequenz drei bis dreizehn und die dritte Untergruppe (F3) von Sequenz vierzehn bis 20 zusammengesetzt (gekennzeichnet mittels rosa Punkten am rechten Rand des MEPAL). Betrachtet man die konservierten AS der Gruppe *Fungi* (siehe Punkt 4, Tabelle 5) mit der Sequenz 2o70 treten zum Teil Übereinstimmungen auf.

Dabei belaufen sich die meisten Übereinstimmungen auf niedrigem Energiebereich, wie beispielsweise die AS an Position 14, 15, 21, 28 und 29 (markiert mit rosa Punkt). Die *Actinobacteria* und *Proteobacteria* werden jeweils in mehrere Gruppen gesplittet. Hierbei überschneiden sich die Gruppen A1, P1 und A2. Diese weisen Ähnlichkeiten bezüglich ihrer Energien auf. Die Gruppe A3 verhält sich auf Energieebene mehr wie die *Metazoa*, daher die Abspaltung von den Bacteria-Gruppen.

Im MEPAL zuerkennen ist weiterhin, dass es Energiebereiche gibt, welche in allen Sequenzen vorkommen. Diese Bereiche sind mit einem braunen Punkt bzw. Rahmen gekennzeichnet. Die im Rahmen stehenden und blau unterlegten AS sind alle unpolar aliphatisch, die AS an den jeweiligen Positionen innerhalb der Sequenzen sind isofunktionelle Austausche. In Abbildung 42 ist dieser eventuell konservierte Bereich ebenfalls mittels eines braunen Rahmens markiert. In dieser Struktur stellt dieser Bereich die Verbindung zwischen den Helices A und der 3.10 Helix dar. Die Gruppe *Metazoa*, welcher die Struktur 2o70 angehört – Sequenz A1L259 – findet sich als eine Gruppe im MEPAL wieder. Die sechs 2o70-Struktur-Sequenzen der Zebrafisch OHCU Decarboxylase sind gelb gerahmt und im Bezug auf die Energie sehr ähnlich bzw. fast identisch.

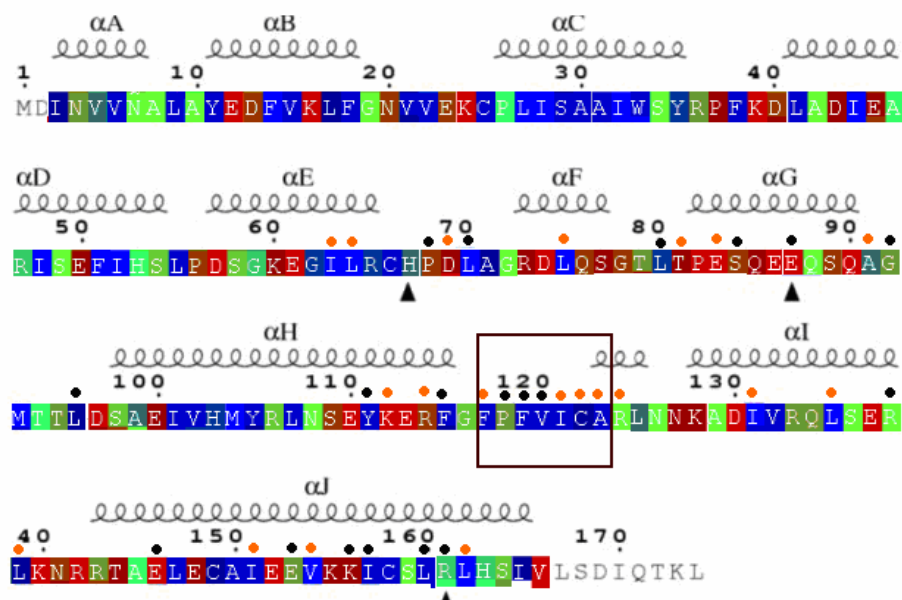


Abbildung 43 2o70 AS Sequenz mit SSE, die Helices sind mit A bis J benannt. Energiefärbung: je **roter** desto **hochenergetischer**, je **blauer** desto **niederenergetischer** ist die AS; Farblich nicht unterlegt AS wurden nicht bewertet. Punktfärbung: schwarz = 100% konserviert innerhalb der Gruppe *Metazoa*, **orange** = isofunktionell.

In der Abbildung 43 ist die 2o70 Sequenz mit Energie-Färbung sowie ihrer dazugehörigen SSE Anordnung zu sehen.

Die Helices A bis J, mit Ausnahme von F und G, werden von meist niederenergetischen AS gebildet. Somit könnte die Theorie, dass Helixstrukturen aufgrund ihrer Stabilität eher eine niedrige Energie aufweisen als instabilere Strukturen, wie beispielsweise Coils, welche sich doch eher durch hochenergetische AS auszeichnen, bestätigen.

Doch wie verlässlich bzw. wie exakt sind die aus Sequenzen gewonnenen Energieprofile wirklich im Vergleich zu Energieprofilen, welche aus Strukturen generiert werden? In der nachfolgenden Abbildung 44 ist ein Ausschnitt eines Test MEPALs der Strukturen (mit allen Ketten je A-F) 2o70, 2o73 und 2o74 zusehen. Der lila Rahmen zeigt die Energievorhersage auf Sequenzebene der Strukturen, während der rosa Rahmen die Energien auf Strukturebene zeigt. Dabei kann festgestellt werden, dass der Großteil der einzelnen AS-Energien der Sequenz-Energievorhersage, trotz weniger Informationen, mit denen der Struktur-Energieprofile übereinstimmt bzw. die Energietendenz anzeigt (je blauer desto niedriger die Energie, je roter desto hochenergetischer). Dies zeigt, dass auch aus wenigen Informationen (reine Sequenz, ohne Struktur) eine relativ gute Aussagekraft von Energieprofilen gewonnen werden kann. Dennoch kann aufgrund von keiner einhundertprozentigen Übereinstimmung beider Energieprofile nur eine Tendenz in der phylogenetischen Analyse bzw. der Untersuchung des MEPALs bezüglich der 80 Sequenzen verwiesen werden.

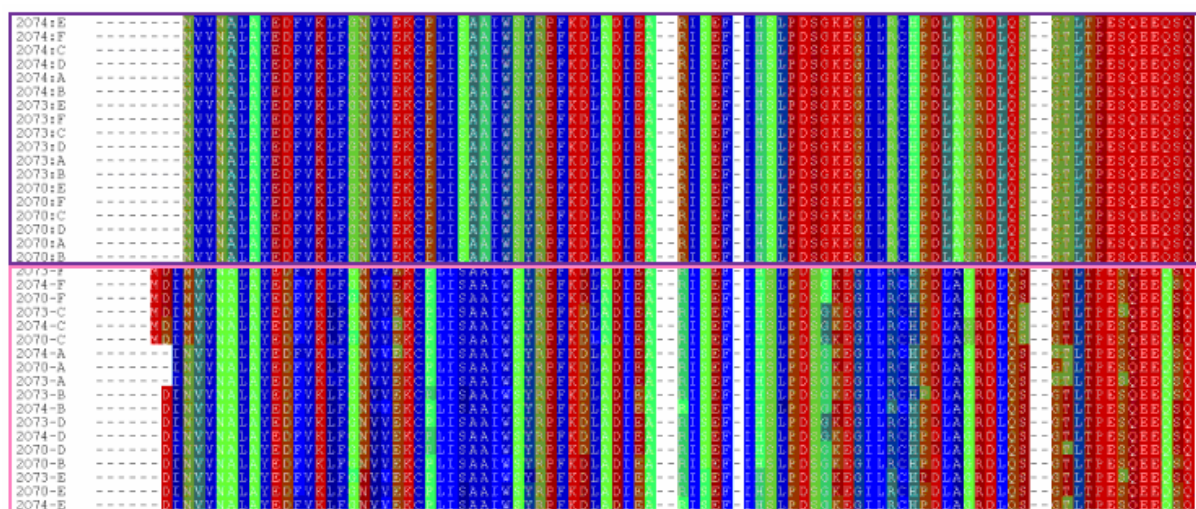


Abbildung 44 MEPAL aus Struktur-Energieprofilen (rosa Rahmen) und Sequenz-Energieprofilen (lila Rahmen). Je blauer = niedriger, je roter = höher ist ihre Energie.

5.3 Erstellung eines phylogenetischen Baumes auf Ebene der Energieprofile von Sequenzen

Der nachfolgende gewurzelte phylogenetische Baum (siehe Abbildung 45) wurde aus den 80 Energieprofilen basierend auf den 80 Sequenzen und aus sechs Energieprofilen basierend auf den Strukturen 2o70 A-F rekonstruiert.

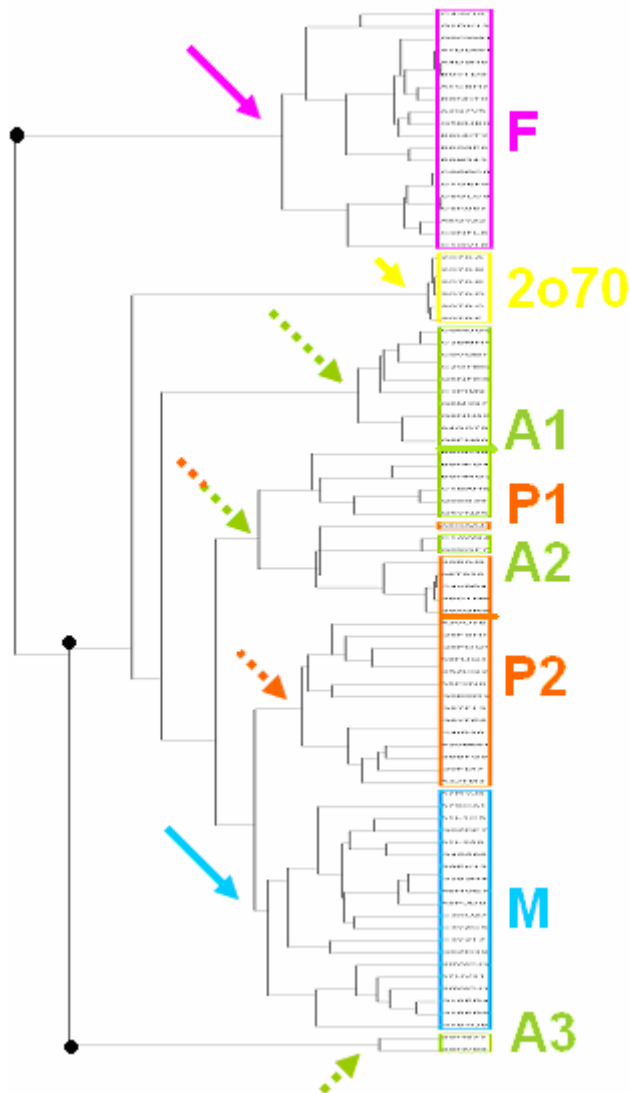


Abbildung 45 Phylogenetischer Baum, basierend auf Energieprofilen der 80 Sequenzen und Struktur 2o70(A-F). Färbung: rosa = *Fungi* (F), grün = *Actinobacteria* (A), blau = *Metazoa* (M), orange = *Proteobacteria* (P).

Jede Organismengruppe wurde farblich entsprechend dem MEPAL (siehe Abbildung 42) markiert. Dabei ist zu erkennen, dass sich die Gruppen wie schon im MEPAL beobachtet heraus bilden. Weiterhin bilden sich drei Hauptgruppen aus dem phylogenetischen Baum hervor (Abtrennung schwarz markierte). Die Fungi bilden die erste Gruppe, die Metazoa, Actinobacteria sowie Proteobacteria bilden eine zweite Gruppe und die Actinobacteria Sequenzen B5H577 & B5HV89 eine dritte Gruppe. F kann als symbolische Outgroup angesehen werden, und somit den phylogenetisch ältesten Ast repräsentieren. Die beiden anderen Gruppen verhalten sich monophyletisch zueinander, d.h. sie besitzen ein und denselben Vorfahren.

Ferner lässt sich erkennen, dass A1 sich in zwei Untergruppen - Sequenzen 1 bis 10 und Sequenzen 11 bis 18 – spaltet (grüne horizontale Linie). Dabei lassen sich diese A1-Untergruppen auf einen letzten gemeinsamen Vorfahren zurückführen, welchen sie sich jedoch mit Angehörigen einer anderen Gruppierung teilen. Die Angehörigen in diesem Falle entsprechen P1, A2 und den P2-Sequenzen 1 bis 6 (orange horizontale Linie). Ebenfalls paraphyletisch verhalten sich die P2-Sequenzen 7 bis 20 zu ihren Sequenzen 1 bis 6. Denn die Sequenzen 7 bis 20 zeigen monophyletische Eigenschaften in Bezug auf M, wobei M die Angehörigen einer anderen Gruppierung repräsentieren, wodurch folglich P2 (1-6) und P2(7-20) ihren letzten gemeinsamen Vorfahren, auf denen sie sich zurückführen lassen, teilen.

6 Vergleich der phylogenetischen Beobachtungen basierend auf Sequenz – und Energieprofilebene

Aus der Abbildung 46 geht die Gemeinsamkeit eines gleichen Vorfahrens hervor, da in beiden Bäumen die Sequenzen einen gemeinsamen Ursprung aufweisen. Weiterhin bilden *Fungi* und *Metazoa* in jedem Baum eine einheitliche Organismengruppe, der keine Sequenz einer anderen Organismengruppe zugehörig ist. Die rosa bzw. blauen Pfeile zeigen den Entspringungspunkt der *Fungi* bzw. *Metazoa*.

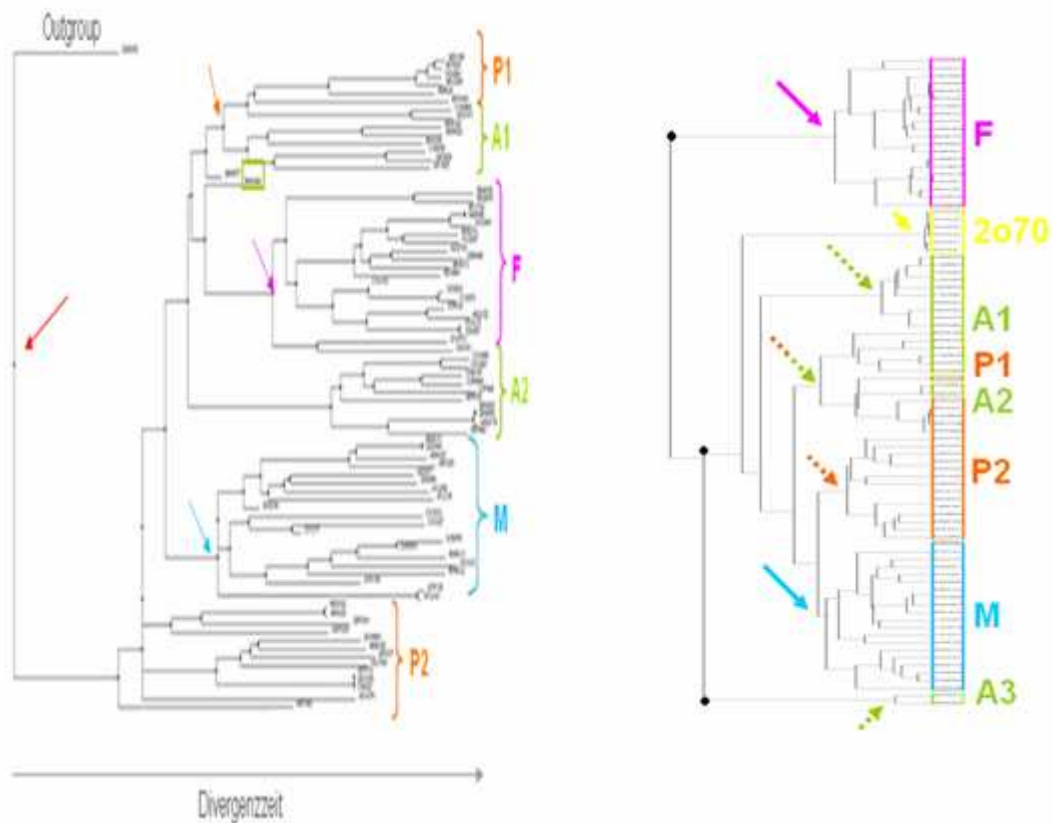


Abbildung 46 Vergleich der phylogenetischen Bäume der OHCU Decarboxylase basierend auf MSA und MEPAL. Links: Dendrogramm MSA basierend, Rechts: Kladogramm MEPAL basierend.

Weiterhin besteht eine Teilung der Proteobacteria in P1 und P2, sowie der Actinobacteria in A1 und A2 bzw. im EP basierenden Baum zusätzlich A3. A3 wird von den Sequenzen B5HV89 und B5H577 im EP-Baum gebildet. Im MSA basierenden Baum hingegen wird B5H577 eher einem gemeinsamen Vorfahr mit P1 und A1 zugeordnet und B5HV89 teilt sich einen Vorfahr mit F. Folglich verhalten sich

diese beiden Sequenzen, B5HV89 und B5H577, laut des MSA basierenden Baumes polyphyletisch zueinander, die beide Mitglieder auf mehr als eine Stammart zurück gehen. Im Gegensatz dazu präsentieren sich beide Sequenzen im EP basierenden Baum als monophyletisch und gehen daher nicht auf mehr als eine Stammart zurück. Die Gruppen A1 und A2 im MSA basierenden Baum sind zueinander paraphyletisch, da sie zwar auf eine letzte gemeinsame Stammform zurückzuführen sind, sich diese jedoch mit den Mitgliedern von P1 teilen. Im EP basierenden Baum zeigen A1 Sequenzen 1-10 und A2 gleiches Verhalten. Die A1-Sequenzen 11-20 verhalten sich zu A2 jedoch monophyletisch.

Dass M, A1, A2, A3, P1 und P2 aus einem gemeinsamen Knoten im EP basierenden Baum hervorgehen und sich somit von F abspalten, kann eventuell an der 100%-igen Konservierung der AS Phenylalanin (siehe Punkt 4.1, Tabelle 5). Diese tritt in allen Gruppen außer in der *Fungi* Gruppe auf und bringt daher eine Gemeinsamkeit zwischen den *Actinobacteria*, *Proteobacteria* und *Metazoa* hervor.

Die unterschiedliche Verteilung der Äste von A, M und P zwischen den MSA basierenden und EP basierenden Bäumen geht womöglich auch auf den höchstwahrscheinlich konservierten Bereich, welcher in Abbildung 41 braun gerahmt wurde, hervor. In diesem Bereich des MEPAL tritt eine große Lücke bei *Fungi* auf, welche von den anderen Gruppen meist mit der AS-Folge F-P-F-V-I-C-V gefüllt wird.

Bei einem Vergleich der phylogenetischen Bäume auf Sequenz- und EP-Ebene muss bedacht werden, dass der auf Sequenzen basierende Baum nur die einzelnen an den jeweiligen Positionen stehenden AS einbezieht und somit auf die Ähnlichkeit zwischen den einzelnen AS-Abfolgen zielt und daraus die evolutionären Distanzen ableitet. Der auf EP basierenden Baum hingegen bezieht zusätzlich die Energien der jeweiligen AS und auch, wenn das Energieprofil aus einer Struktur abgeleitet wurde, strukturelle Eigenschaften in die evolutionären Kalkulationen ein und könnte bzw. sollte daher genauere Aussagen über die Evolution dieses Enzyms liefern. Daher liefert der EP basierende phylogenetische Baum genauere Aussagen über die evolutionäre Entwicklung der Enzymfamilie der 2-oxo-4-Hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase.

7 Ausblick

Um einen noch exakteren Überblick und genauere phylogenetische Aussagen auf Grundlage der Sequenzen über die Familie der OHCUs Decarboxylasen zu erhalten bzw. treffen zu können, kann man in einer weiteren Arbeit alle ca. 230 Sequenzen der Familie alignen. Basierend auf einem WebLogo aus diesem multiplen Alignment würde man für alle Sequenzen den Grad der Konservierung aller Aminosäuren an den jeweiligen Positionen erkennen. Dieses Logo könnte somit die Annahmen über die an der Katalyse mutmaßlich beteiligten Aminosäuren von Ramazzina et al. als auch Kim et al. unterstützen. Des Weiteren bietet sich an, einen phylogenetischen Baum, resultierend aus dem zuvor erstellten multiplen Sequenz Alignment, zu rekonstruieren. Ein Baum, welcher auf dem eher einfachen und unkomplizierten Berechnungsverfahren des Neighbor Joinings (NJ) beruht, eignet sich für große Datensätze wie im Falle der OHCUs Decarboxylasen. Diese Methode liefert mehr oder weniger erste Abschätzungen zur Phylogenie. Um akkurate und signifikante Aussagen zu treffen, sollte man ein aufwendigeres Verfahren, wie beispielsweise Maximum Parsimony (MP), wählen. Für umfangreiche und sehr signifikante Aussagen auf Ebene der Energieprofile der Familie OHCUs Decarboxylase, könnte ein Algorithmus gebildet werden, der alle Energieprofile der Familie berechnet und diese Energieprofile aligniert. Somit könnten bessere Vergleiche zu den Sequenzen bzw. dem multiplen Sequenz Alignment durchgeführt werden und daraus eventuell eindeutigerere Aussagen zur Evolution dieser Familie getroffen werden. Ferner wäre es von großem Nutzen zur Erstellung der Energieprofile, allen Sequenzen eine dreidimensionale Struktur zuzuordnen zu können bzw. eine Tertiär- bzw. Quartärstruktur zu bestimmen, da bislang lediglich 2 Strukturen aufgeklärt werden konnten. Dies könnte mit Hilfe der SwissProt DB gelingen, in dem jeweils zwei bis drei Vertreter jeder Organismengruppe auf die Struktur 2q37 und 2o70 modelliert werden um zu sehen. Sehr interessant wäre es zudem, wenn eine Wiederaufnahme der Funktion der (menschlichen) OHCUs Decarboxylase erzeugt werden könnte. Dazu muss das notwendige Gen sozusagen „angeknipst“ werden. Doch leider ist dieser biologische Vorgang noch unklar. Eine erfolgreiche Nutzung der menschlichen OHCUs Decarboxylase könnte den Menschen und auch anderen Säugetieren, die von einem „ausknipsen“ des OHCUs Decarboxylase Gens betroffen sind, von Volkskrankheiten wie der Gicht heilen.

Anhang

Tabelle 9 Überblick über Chemisch-Physikalische-, toxikologische- und ökotoxikologische Eigenschaften von Allantoin [Quelle: Merck: Sicherheitsdatenblatt gemäß Richtlinie 91/155/EWG. Merck KGaA Darmstadt].

<u>Allgemeine Angaben</u>	
Synonyme	5-Ureido-hydantoin Glyoxylsäurediureid
CAS-Nr.: 97-59-6	EG-Nr.: 202-592-8
Summenformel	C ₄ H ₆ N ₄ O ₃
<u>Chemisch-Physikalische-Eigenschaften</u>	
Molmasse	158,12 g mol ⁻¹
Form	fest
Farbe	weiß
Geruch	geruchlos
pH-Wert (20°C) bei 5 g l ⁻¹ H ₂ O	4,5 – 6,0
Thermische Zersetzung	225-236 °C
Dichte	~ 1,7 g cm ³
Schüttdichte	~ 800 kg m ³
Löslichkeit in Wasser (25°C)	5,7 g l ⁻¹
in Wasser (75°C)	40 g l ⁻¹
in Ethanol (25°C)	fast unlöslich
<u>Toxikologie</u>	
LD ₅₀ (oral, Ratte)	> 5000 mg kg ⁻¹
<u>Ökotoxikologie</u>	
Biologische Abbaubarkeit	gut, > 70%
Fischtoxizität: LC ₅₀ Brachydanio rerio	> 5000 mg l ⁻¹
Bakterientoxizität: EC ₀ <i>Pseudomonas putida</i>	> 10 g l ⁻¹
CSB	240 mg g ⁻¹
BSB ₅	63,8 mg g ⁻¹

Literaturverzeichnis

Bücher

- [1] Fuchs, Georg; Schlegel, Hans G.: Allgemeine Mikrobiologie. Stuttgart: Thieme Verlag. 8. Auflage, 2007
- [2] Geadecke, Nicola: Biowissenschaftlich recherchieren – Über den Einsatz von Datenbanken und anderen Ressourcen in der Bioinformatik. Berlin: Birkhäuser Verlag, 2007
- [3] Rehm, Hubert; Hammar, Friederike: Biochemie light. Frankfurt a.M.: Verlag Harri Deutsch, 3. korrigierte Auflage, 2005
- [4] Horton, Robert H.; Moran, Laurence A.; Scrimgeour, Gray K.; Perry, Marc D.; Rawn, David J.: Biochemie. München: Pearson Studium, 4. aktualisierte Auflage, 2008
- [11] Hansen, Andrea: Bioinformatik – Ein Leitfaden für Naturwissenschaftler. Berlin: Birkhäuser Verlag, 2. überarbeitete und erweiterte Auflage, 2004
- [13] Merkl, Rainer; Waack, Stephan: Bioinformatik Interativ – Grundlagen, Algorithmen, Anwendungen. Weinheim: Wiley-VCH Verlag, 2. erweiterte und neubearbeitete Auflage, 2009
- [14] Storch, V.; Welsch, U.; Wink, M.: Evolutionsbiologie. Berlin: Springer Verlag, 2001
- [18] Koolman, Jan; Röhm, Klaus-Heinrich: Taschenatlas der Biochemie.- 3. vollständig überarbeitete und erweiterte Auflage – Stuttgart: Georg Thieme Verlag, 2003

Vorlesungsskripte/ Bachelorarbeiten/ Dissertationen

- [5] Heinke, Florian: Energieprofilbasierende Analysemethoden von Proteinfamilien. - Bachelorarbeit, Hochschule Mittweida - Mittweida, 2010
- [6] Dressel, Frank: Sequenz, Energie, Struktur – Untersuchungen zur Beziehung zwischen Primär- und Tertiärstruktur in globulären und Membran-Proteinen. Dissertation, Technische Universität Dresden – Dresden, 2008
- [7] Dr. Schwenzer, B.: Skript zur Vorlesung Molekularbiologie/Gentechnik. - Hochschule Mittweida, Mittweida 2009
- [15] Prof. Dr. rer. nat Dirk Labudde: Skript zur Vorlesung Biodatenbanken.- Hochschule Mittweida - Mittweida 2009

Paper

- [9] Finn, Robert D. at all: The Pfam Protein Families Database. Nucleic Acid Research, Vol. 36, 2008
- [10] Sonnhammer, Erik L.L. at all: Pfam: A Comprehensive Database of Protein Domain Families Based on Seed Alignments. PROTEINS:Structure, Function and Genetics, 1997
- [16] Ramazzina, Illeana; Berni, Rudolfo; Zanotti, Guiseppa: The Structure of 2-Oxo-4-hydroxy-4-carboxy-5-ureidoimidazoline Decarboxylase Provides Insights into the Mechanism of Uric Acid Degradation. The Journal of Biological Chemistry Vol. 282, NO 25, June 22, 2007
- [17] Kim, Kwangsoo; Park, Jineso; Rhee, Sangkee: Structural and Functional Basis for (S)-Allantoin Formation in the Ureide Pathway. The Journal of Biological Chemistry Vol. 282, NO 32, August 10, 2007
- [21] Ramazzina, Ileana; Folli, Claudia; Secchi, Andrea: Completing the uric acid degradation pathway through phylogenetic comparison of whole genomes. Nature Chemical Biology Vol. 2, NO 3, March 2006
- [22] Tipton, Peter A.: Urate to allantoin, specifically (S)-allantoin. Nature Chemical Biology Vol. 2, NO 3, March 2006

Internet

- [8] <http://www.chemgapedia.de>
- [12] <http://www.pfam.sanger.ac.uk>
- [19] <http://www.ebi.ac.uk>
- [20] <http://www.pdb.org>
- [23] <http://www.jalview.org>
- [24] <http://weblogo.berkeley.edu>
- [25] <http://bioservices.hs-mittweida.de>